EWARS: Practical aspects of risk mapping

1. Data structure and contents

The current EWARS model, allows two different approaches of risk mapping; i) ***MODEL 1***: using records derived from aggregated spatial units e.g. disease cases and alarm indicators data aggregated at district, sub-district, village, neighborhood or household levels and, ii) ***MODEL 2***: using geo-tagged disease cases (point) data. This case reporting includes the longitude and latitude coordinates of where the unit of analysis is being reported within a defined area, e.g. hospital, laboratories, households, schools...etc.

The risk mapping analysis will require historical data to provide users with an overview of "hotspot" patterns over time. This will show how some 'hotspots' have changed in past weeks (or years) within a defined area, which will guide future vector control plans. Dataset uploaded for risk mapping (tab) should contain historical information including information on most recent disease and alarm indicators. The longer the historical period the better but it is recommended to have at least 4 weeks of most recent disease and alarm indicator records (i.e. data from current week + previous weeks).

Please note that, for ***MODEL 1***, only one level of data is accepted. For instance, you may process data at district level or village level but not both at the same time.

Below are some examples of data structures for each MODEL:

*Figure 1a, example of data for **MODEL 1** – using data aggregated at district level. (NB. 'District' here represents the ID of the smallest geographical unit; this will be useful for comparing spatial units (districts) with each other by the central level).*

year	district	population	week	weekly_hospitalised_cases	rhdailymean	rainsum	meantemperature
2008	1	1141000	44	0	67.065714	0	18.6814285714286008
2008	1	1141000	45	0	31.841429	0	20.5735714285714018
2008	1	1141000	46	0	30.802857	0	22.3999999999999986
2008	1	1141000	47	0	25.982857	0	15.8285714285714008
2008	1	1141000	48	0	42.721429	0	20.814285714285699
2008	1	1141000	49	0	39.172857	0	16.0407142857143015
2008	1	1141000	50	0	34.857143	0	17.092857142857099
2008	1	1141000	51	0	44.661429	0	18.6714285714285992
2008	1	1141000	52	0	47.072857	0	16.2571428571428989
2009	1	1138000	1	0	50.206667	0	18.5214285714286007
2009	1	1138000	2	0	53.922857	0	14.3499999999999996
2009	1	1138000	3	0	55.5	0	19.8071428571428996
2009	1	1138000	4	0	49.714286	0	16.9928571428571011
2009	1	1138000	5	2	63	0	17.9285714285713986
2009	1	1138000	6	1	52	0	21.6000000000000014
2009	1	1138000	7	0	47	0	19.2071428571429017
2009	1	1138000	8	0	75.637143	0	21.5785714285714008
2009	1	1138000	9	0	42.612857	0	19.7928571428570983
2009	1	1138000	10	2	56.751667	0	18.1999999999999993

Figure 1b, example of data for ***MODEL 1****– using data aggregated at smaller spatial unit, e.g. village level within a defined area. (NB. 'Village' here represents the ID of the smallest geographical unit, this will be useful for comparing spatial units (villages) with each other by the 'District' level).*

	year	district	village	population	week	weekly_hospitalised_cases	rhdailymean	rainsum	meantemperature
298	2013	1	1	1124000	38	7	88.9	250	23.8
299	2013	1	1	1124000	39	18	64.1	3	26.9
300	2013	1	1	1124000	40	20	72.7	63	26.8
301	2013	1	1	1124000	41	31	58.6	0	23.7
302	2013	1	1	1124000	42	21	81.7	30	22.9
303	2013	1	1	1124000	43	12	67.7	1.2	20.5
304	2013	1	1	1124000	44	12	70.1	0	24.2
305	2013	1	1	1124000	45	14	68.2	52.8	19.3
306	2013	1	1	1124000	46	4	.	.	.
307	2013	1	1	1124000	47	3	86	6.4	7.5
308	2013	1	1	1124000	48	4	70.6	11.4	10.7
309	2013	1	1	1124000	49	2	60.7	0	18.9
310	2013	1	1	1124000	50	1	72.7	0	11.5
311	2013	1	1	1124000	51	0	55.4	0	17.5
312	2013	1	1	1124000	52	0	72	31.4	12.3
313	2008	1	2	125000	1	0	34.5	0	14.6
314	2008	1	2	125000	2	0	43.7	0	19.9
315	2008	1	2	125000	3	0	76.1	0	11.3
316	2008	1	2	125000	4	0	73.6	0	14
317	2008	1	2	125000	5	0	46.1	0	18
318	2008	1	2	125000	6	0	35.1	0	23
319	2008	1	2	125000	7	0	54.7	0	22.8
320	2008	1	2	125000	8	0	52.4	0	23.3
321	2008	1	2	125000	9	0	47.3	0	21.5
322	2008	1	2	125000	10	0	38	0	18.7
323	2008	1	2	125000	11	0	37.4	0	23.9

*Figure 2a, example of point data for **MODEL 2** – using point data. The 'X' and 'Y' columns represents the longitude and latitude coordinates of data source points. In this figure, we present example of geo-tagged data from household level within a defined geographical area (e.g. district or region).*

	district	year	x	y	week	rainsum	mintemperature	meantemperature	maxtemperature
49	15	2020	-104.5209	24.08692	1	34.4	13.3999996185302734	22	30.60000038146972656
50	15	2020	-104.5182	23.29564	1	39.6	15.1999998092651367	22.7000007629394531	30.20000076293945313
51	15	2020	-104.3942	25.24401	1	14.7	18.3999996185302734	27.8999996185302734	37.40000152587890625
52	15	2020	-104.3085	23.12993	1	41.2	15.1999998092651367	23.2000007629394531	31.20000076293945313
53	15	2020	-104.2981	24.94559	1	20.9	15.8999996185302734	25.2999992370605469	34.70000076293945313
54	15	2020	-104.203	22.83854	1	47.3	16.7000007629394531	24.2999992370605469	32
55	15	2020	-104.1694	25.21993	1	14.7	18.3999996185302734	27.8999996185302734	37.40000152587890625
56	15	2020	-104.0696	24.44785	1	27.3	14.8000001907348633	23.5	32.20000076293945313
57	15	2020	-104.0424	25.14582	1	14.7	18.3999996185302734	27.8999996185302734	37.40000152587890625
58	15	2020	-104.0026	26.7514	1	11.9	19.7000007629394531	29	38.40000152587890625
59	15	2020	-103.9881	24.35829	1	25.3	13.8000001907348633	23	32.20000076293945313
60	15	2020	-103.8797	26.1749	1	8.1	19.7000007629394531	29	38.29999923706054688
61	15	2020	-103.4844	25.34185	1	9.4	18.8999996185302734	27.7000007629394531	36.59999847412109375
62	15	2020	-103.3445	26.28322	1	7.9	19.5	28.6000003814697266	37.79999923706054688
63	15	2020	-103.0263	24.50444	1	14.4	17.7999992370605469	26.7999992370605469	35.90000152587890625
64	15	2020	-102.9134	24.56134	1	12.4	17.7000007629394531	26.6000003814697266	35.59999847412109375
65	15	2020	-102.5786	24.62329	1	12.4	17.7000007629394531	26.6000003814697266	35.59999847412109375
66	15	2020	-102.4931	24.47041	1	15.9	15	24.7999992370605469	34.59999847412109375
67	15	2020	-107.131	25.49996	2	115	22.7000007629394531	28.2999992370605469	34
68	15	2020	-106.9731	25.07157	2	129.8	17.5	22.8999996185302734	28.39999961853027344
69	15	2020	-106.9435	25.42512	2	129.8	17.5	22.8999996185302734	28.39999961853027344
70	15	2020	-106.7223	25.18269	2	129.8	17.5	22.8999996185302734	28.39999961853027344
71	15	2020	-106.6578	25.49998	2	129.8	17.5	22.8999996185302734	28.39999961853027344
72	15	2020	-106.5509	24.60299	2	99.3	22.5	27.6000003814697266	32.79999923706054688
73	15	2020	-106.4897	24.73073	2	104.5	17.7999992370605469	23.2000007629394531	28.70000076293945313
74	15	2020	-106.4896	25.52074	2	128.4	11.6999998092651367	18.1000003814697266	24.60000038146972656
75	15	2020	-106.4686	24.65816	2	104.5	17.7999992370605469	23.2000007629394531	28.70000076293945313
76	15	2020	-106.4049	25.14413	2	104.8	11.5	17.6000003814697266	23.70000076293945313
77	15	2020	-106.3989	25.27381	2	104.8	11.5	17.6000003814697266	23.70000076293945313
78	15	2020	-106.2937	25.85201	2	128.4	11.6999998092651367	18.1000003814697266	24.60000038146972656
79	15	2020	-106.2876	24.47268	2	110.7	17.7999992370605469	22.7999992370605469	27.89999961853027344
80	15	2020	-106.2843	25.26222	2	104.8	11.5	17.6000003814697266	23.70000076293945313

*Figure 2b, example of point data for **MODEL 2** – using point data. The 'X' and 'Y' columns represents the longitude and latitude coordinates of data source points. In this figure, we present example of geo-tagged data from local Hospital-level (where cases are reported from) within a defined geographical area (e.g. district or region).*

	district	year	x	y	week	rainsum	mintemperature	meantemperature	maxtemperature
1	15	2020	-106.9335	24.85958	1	44.5	22.4	28.6	34.8
2	15	2020	-106.431	26.11964	1	44	11.1	19.7	28.3
3	15	2020	-106.4264	25.66018	1	47.3	11.4	19.5	27.6
4	15	2020	-106.3933	24.73316	1	50.8	18.2	24.8	31.4
5	15	2020	-106.2781	26.0126	1	44	11.1	19.7	28.3
6	15	2020	-106.2143	25.54838	1	47.3	11.4	19.5	27.6
7	15	2020	-106.1312	26.74512	1	32.2	15.1	23.5	31.9
8	15	2020	-106.9335	24.85958	2	115	22.7	28.3	34
9	15	2020	-106.431	26.11964	2	129.8	17.5	22.9	28.4
10	15	2020	-106.4264	25.66018	2	129.8	17.5	22.9	28.4
11	15	2020	-106.3933	24.73316	2	129.8	17.5	22.9	28.4
12	15	2020	-106.2781	26.0126	2	129.8	17.5	22.9	28.4
13	15	2020	-106.2143	25.54838	2	99.3	22.5	27.6	32.8
14	15	2020	-106.1312	26.74512	2	104.5	17.8	23.2	28.7
15	15	2020	-106.9335	24.85958	3	284.6	22.6	27.6	32.7
16	15	2020	-106.431	26.11964	3	261.2	19.1	24.2	29.3
17	15	2020	-106.4264	25.66018	3	284.6	22.6	27.6	32.7
18	15	2020	-106.3933	24.73316	3	280.2	17.8	22.7	27.6
19	15	2020	-106.2781	26.0126	3	280.2	17.8	22.7	27.6
20	15	2020	-106.2143	25.54838	3	263.4	13.3	18.7	24.2
21	15	2020	-106.1312	26.74512	3	280.2	17.8	22.7	27.6
22	15	2020	-106.9335	24.85958	4	179.2	16.9	22.1	27.3
23	15	2020	-106.431	26.11964	4	179.2	16.9	22.1	27.3
24	15	2020	-106.4264	25.66018	4	162	12	17.8	23.6
25	15	2020	-106.3933	24.73316	4	179.2	16.9	22.1	27.3
26	15	2020	-106.2781	26.0126	4	260.7	17.1	21.9	26.8
27	15	2020	-106.2143	25.54838	4	231.1	17.2	22.4	27.6
28	15	2020	-106.1312	26.74512	4	183.3	10.5	16.4	22.4
29	15	2020	-106.9335	24.85958	5	45.7	19.5	26.3	33.1
30	15	2020	-106.431	26.11964	5	51.1	20	26	32
31	15	2020	-106.4264	25.66018	5	52.5	15.6	21.5	27.4
32	15	2020	-106.3933	24.73316	5	43.1	10.1	16.6	23.1
33	15	2020	-106.2781	26.0126	5	52.5	15.6	21.5	27.4
34	15	2020	-106.2143	25.54838	5	52.5	15.6	21.5	27.4

If the hospital names are known, then geocode can be obtained easily from google earth:

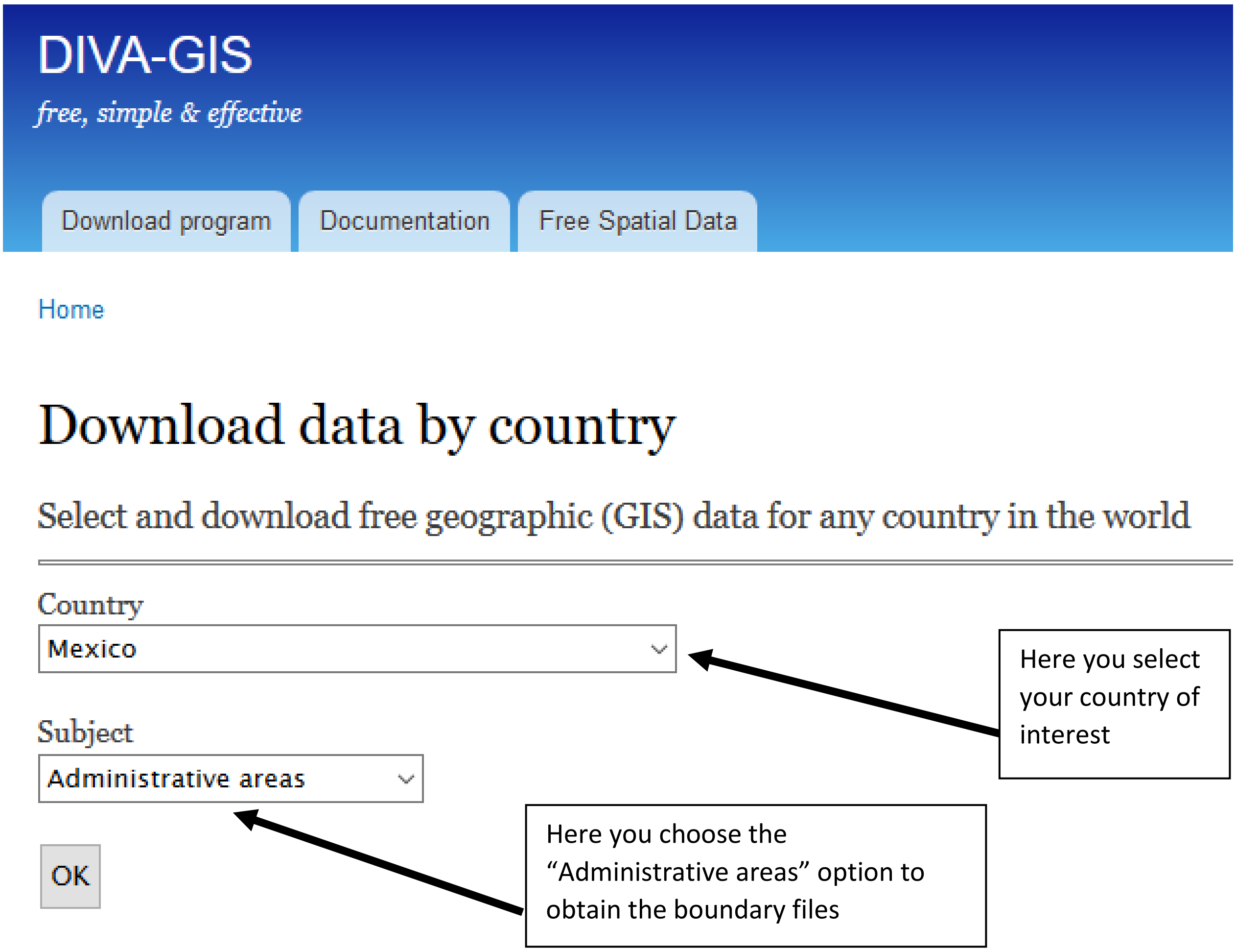
Use Google Earth to look up a place's longitude (east-west position) and latitude (north-south position).

1. *On your computer, open Google Earth.*
2. *On the left, click Search .*
3. *Search for a place. On the right, you'll see a box, or Knowledge Card, with some information about the place.*
4. *In the top left corner of the Knowledge Card, click the Up arrow.*
5. *In the box that appears, you'll see the longitude and latitude.*

2. Boundary files and merging disease information

Before you could run the Risk mapping analysis under EWARS, users need to have their surveillance data uniquely attached to the spatial boundaries of the administrative unit (e.g. district, villages within districts, or other sub-units). We have provided some links for open-access sources to obtain your local boundary files. You may obtain your boundary files from these links. Below, is an illustration of this process using one of these links:

Figure 3, accessing open-access spatial boundary files



There are usually several files generated from this option, however, you would be interested in selecting and downloading four files with the below extensions:

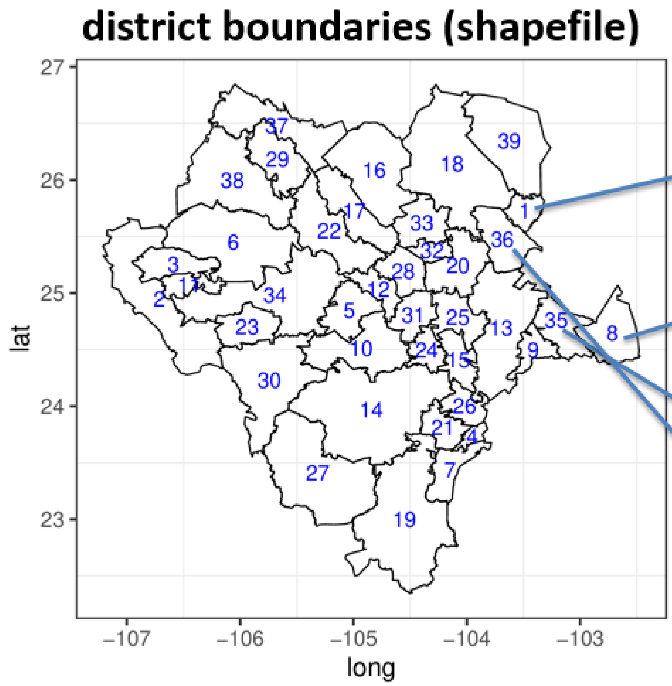
Shape_file_Demo.dbf
Shape_file_Demo.prj
Shape_file_Demo.shp
Shape_file_Demo.shx

The boundary file with extension ".dbf" can be used to identify the ID given for your administrative unit (e.g. district within the country, villages within districts, or other smaller sub-units within a defined area). The ID of the administrative units as provided from the ".dbf" files should then be used in the surveillance disease data to unify the administrative units registered in the surveillance data – this way you are said to link the boundary files of your selected region with the corresponding surveillance data. Figure 4 below shows an example of ".dbf" file showing different IDs given for different districts (from Mexico). The same IDs should be used for the corresponding districts in the surveillance data. The same can applies if you have "villages ID" for one district – then the villages ID obtained from the boundary file (".dbf") should be linked to the corresponding villages in the surveillance data for a defined district. Figure 5, illustrate how the linkage can be viewed in principle.

Figure 4, linking boundary files to surveillance data.

	CVE_ENT	CVE_MUN	NOM_MUN	OID_1	cov_	cov_id	district
0	10	007	G\xf3mez Palacio	343	343	344	1
1	10	034	Tamazula	344	344	345	2
2	10	037	Topia	345	345	346	3
3	10	038	Vicente Guerrero	346	346	347	4
4	10	039	Nuevo Ideal	347	347	348	5
5	10	035	Tepehuanes	348	348	349	6
6	10	033	S\xfachil	349	349	350	7
7	10	027	San Juan de Guadalupe	350	350	351	8
8	10	031	Santa Clara	351	351	352	9
9	10	001	Canatl\xe1n	352	352	353	10
10	10	002	Canelas	353	353	354	11
11	10	003	Coneto de Comonfort	354	354	355	12
12	10	004	Cuencam\xe9	355	355	356	13
13	10	005	Durango	356	356	357	14
14	10	008	Guadalupe Victoria	357	357	358	15

Figure 5, linking boundary files to surveillance data



Surveillance data

district	year	pop	cases	temp
1	2009	1138000	3	31.6
7	2009	137000	2	28.8
8	2009	124000	1	29
19	2009	679000	15	
35	2009	700000	0	42.2
36	2009	424000	0	19.4
39	2009	218000	0	28.4

Further instructions are available in the below link using the QGIS system:

https://gis.stackexchange.com/questions/75563/renaming-attributes-fields-in-shapefile-attribute-table-using-qgis

Once this process is complete, your spatial and surveillance data will be automatically identified once you upload the data + boundary files under the Dashboard, as you can see in the following section.

3. Running the Risk mapping tool in EWARS

- Dashboard I will remain as a platform for calibrating the temporal prediction model and generate sensitivity/ PPV to assess the performance of the prediction.
- Dashboard II is the platform for entering prospective disease and alarm information on a weekly basis and generate an alarm signal predicting outbreak or no-outbreak within the prediction period.
- Dashboard II now accommodates the risk mapping analysis. In principle, users can start with the temporal prediction (predicting the time of outbreak) then predicting the space (geo-location of a possible outbreak) using a historical and more recent (i.e. current week) disease and alarm information.
- An outline of the risk mapping processes is displayed hereunder.
- If your surveillance data does not contain information (about disease cases and alarm indicators) for the entire geo-locations (i.e. districts) of the country, only districts with information will be displayed and the rest will be displayed as missing! Please note that better risk mapping prediction will be achieved when information from all districts are provided since the prediction is adjusted for the outbreak situation in neighboring districts.
- **PLEASE NOTE THAT, IF YOUR SURVEILLANCE DATA OF THE CORRESPONDING DISTRICT DOES NOT CONTAIN EPIDEMIOLOGICAL WEEKS THAT FOLLOW IN SUBSEQUENT ORDER (WEEK 1, 2, 3, 4…53, INSTEAD OF RANDOM WEEKLY RECORDS LIKE WEEK 3, 6, 34, 45), THEN THE ANALYSIS FOR THIS CORRESPONDING DISTRICT WILL NOT BE DISPLAYED!**

Figure 6, Location and explanation of Risk Mapping options under EWARS

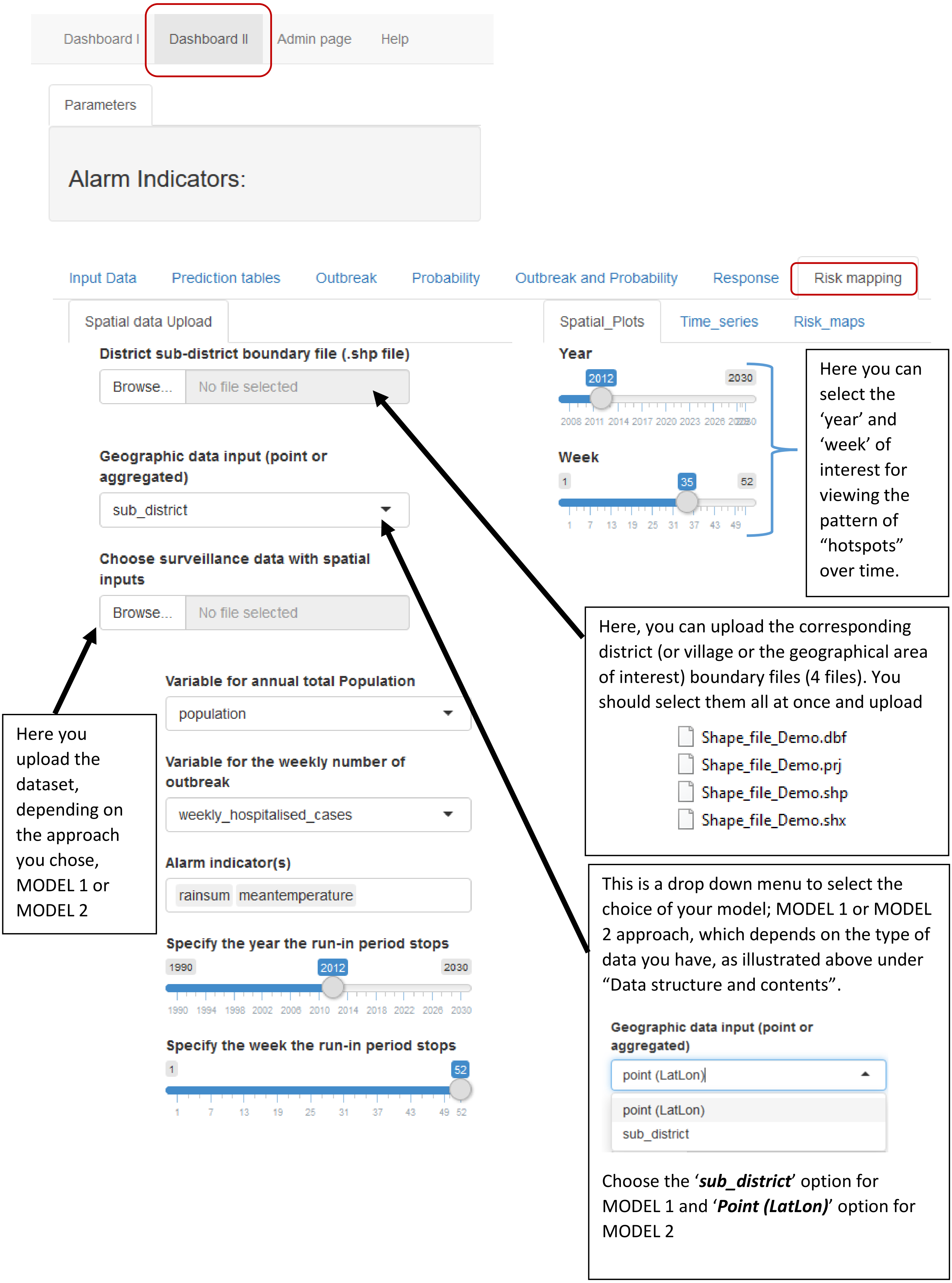
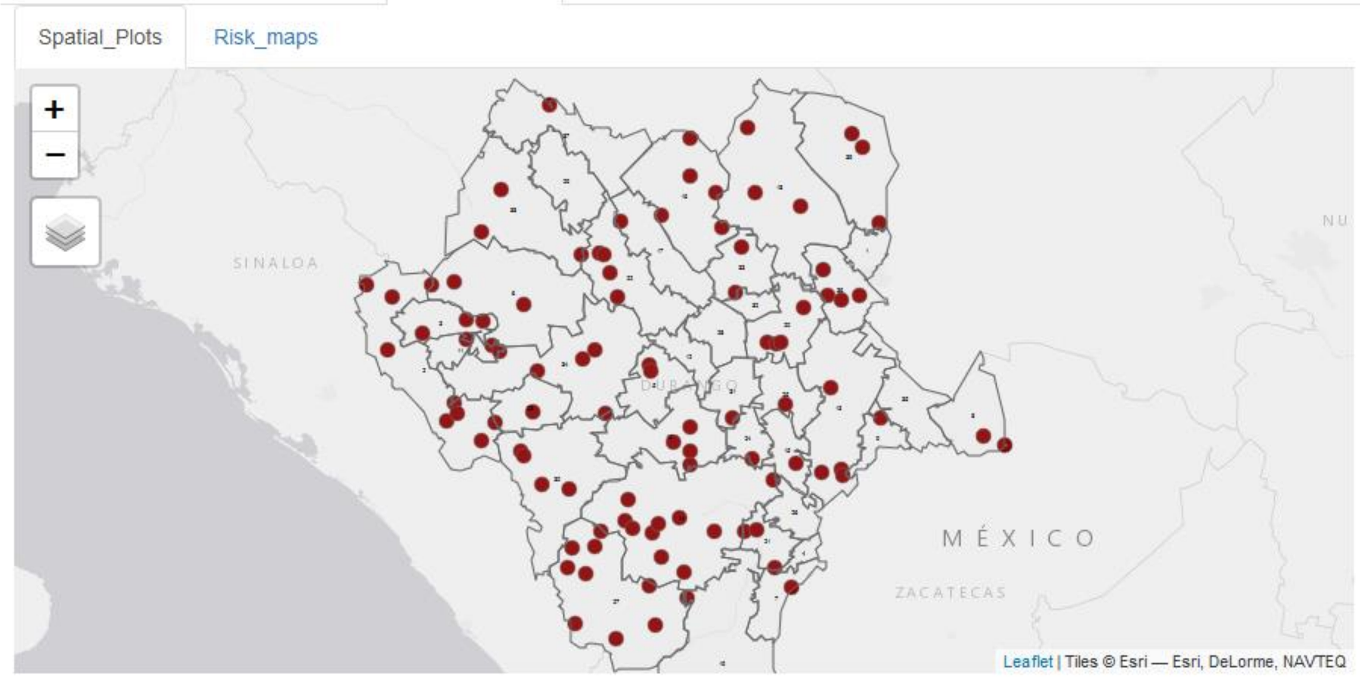


Figure 7a, Example of risk mapping analysis to show "Hotspots" within a defined area (district) using a collection of variables including disease cases (geo-tagged information), alarm indicators and cases at neighboring districts.



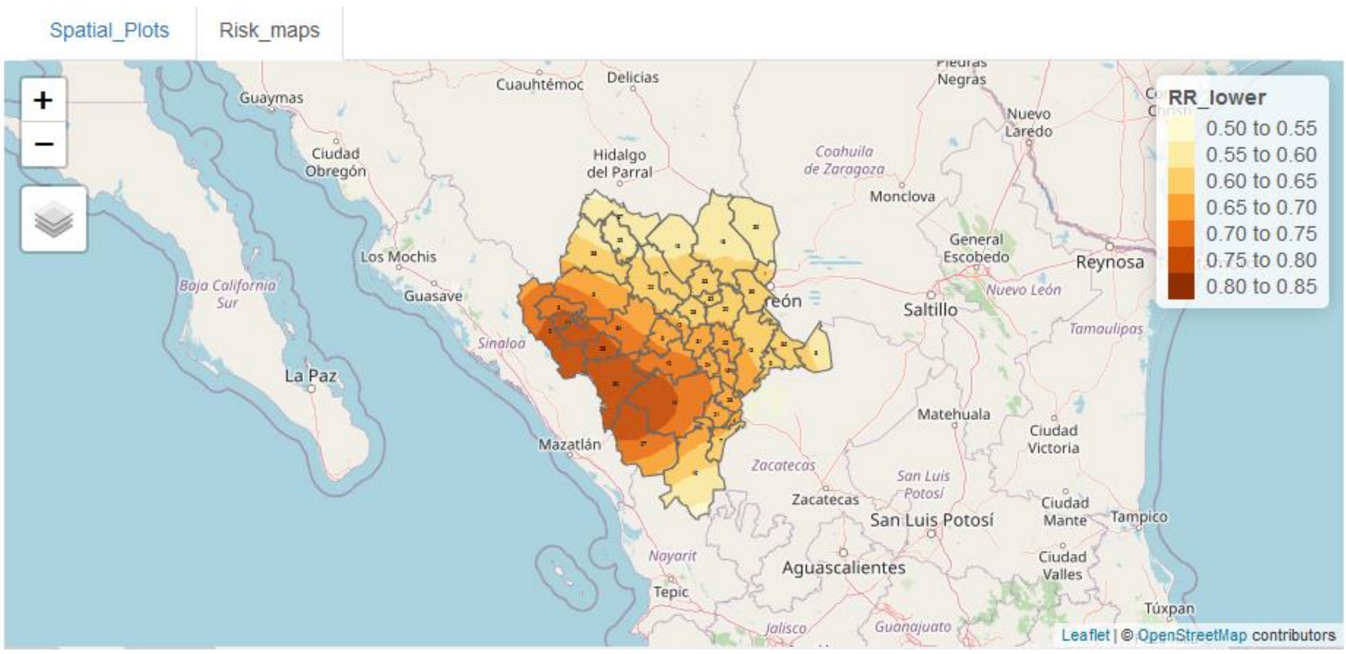


Figure 7b, Example of risk mapping analysis to show "aggregated" data within a spatial unit. Here data are aggregated at 'District' level using a collection of variables including disease cases, alarm indicators. The 'spatial plots' show details of the pattern of disease cases, expected cases (based on the population), the population in each district and, the intensity of the alarm indicators in each district.

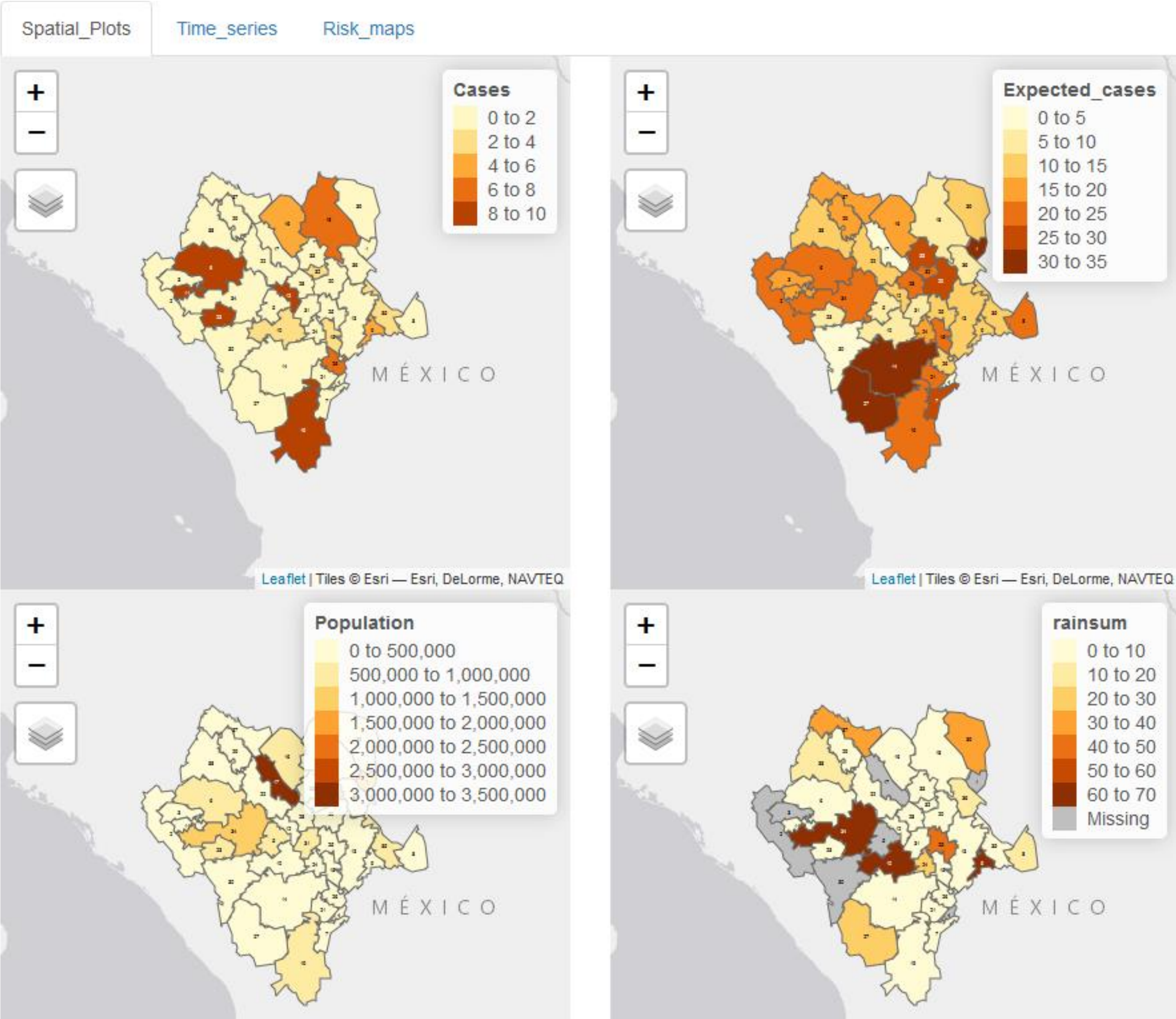
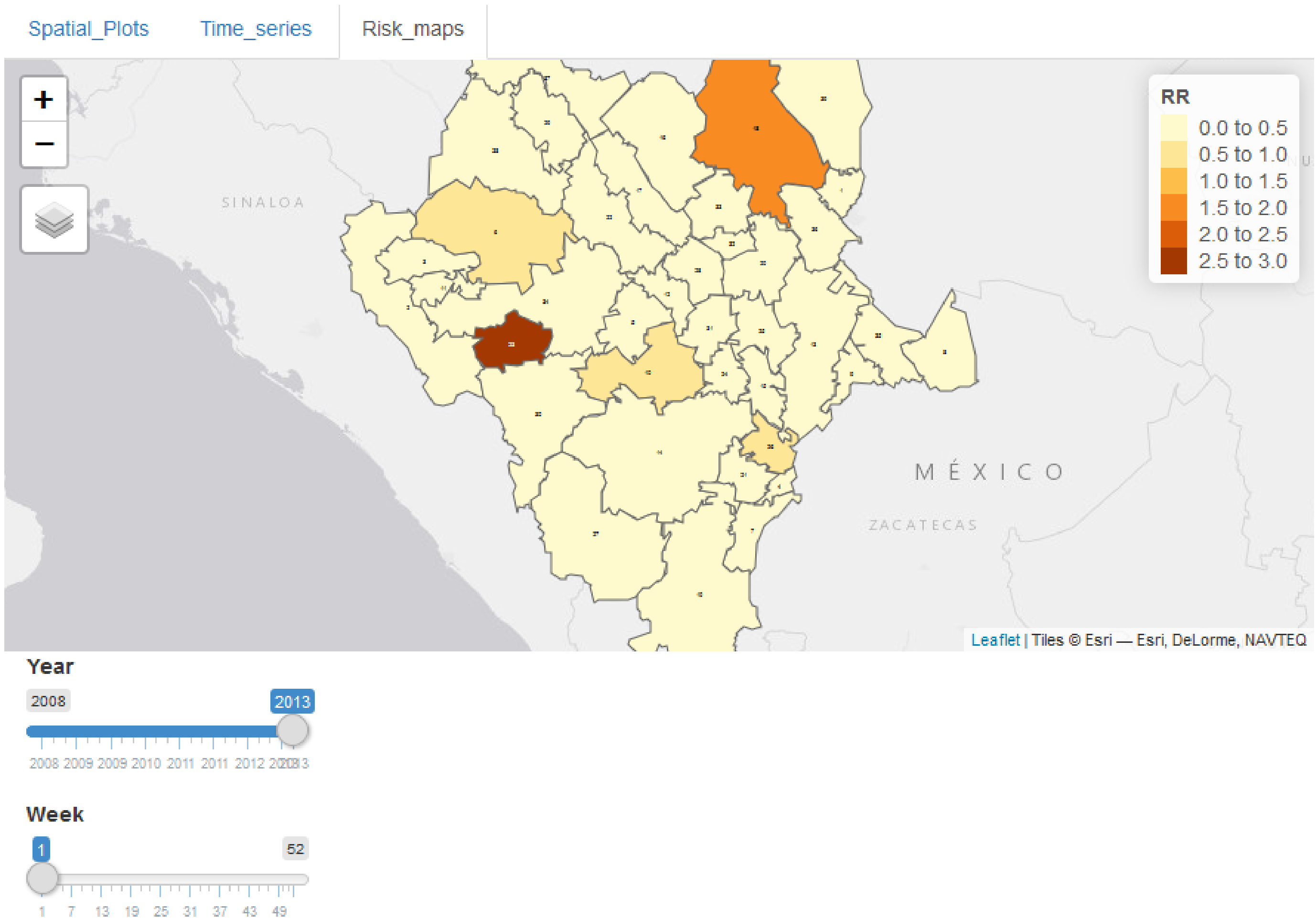


Figure 7c, Example of risk mapping analysis to show "aggregated" data within a spatial unit. Here data are aggregated at 'District' level (by the central level and by the district team to compare the risk in different districts) using a collection of variables including disease cases, alarm indicators. The 'risk maps' show the risk of disease observed in relation to the expected number of cases (based on the disease pattern in the target population, neighboring districts and the alarm indicators).



4. Supporting links

Users can generate their boundary files of the corresponding geographical are of interest by using the open access links, provided hereunder:

- https://spatialdata.dhsprogram.com/boundaries/#view=table&countryId=AF
- https://www.diva-gis.org/gdata
- https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries