



Module INM433 – Visual Analytics

Practical 06

Data import and export in V-Analytics

given by

prof. Gennady Andrienko and

prof. Natalia Andrienko

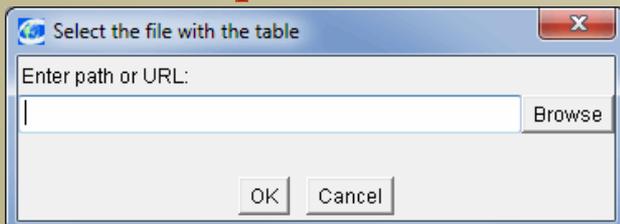
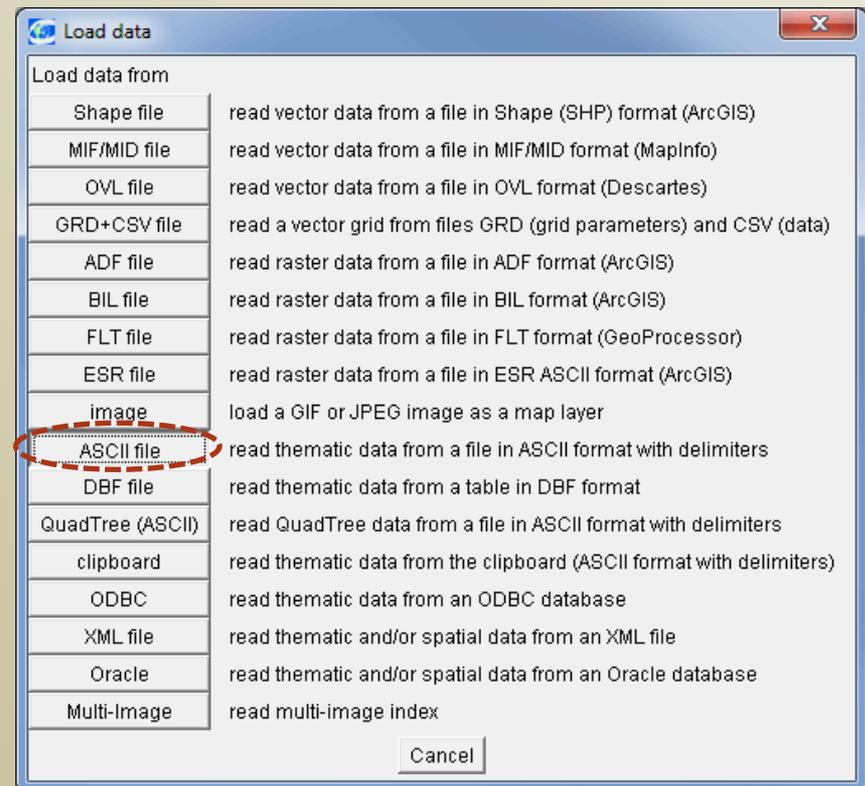


Example of loading and
indexing spatial events from a
text file



1) Loading a table from a text file

- Supported format: text with separators (comma, semicolon, tab, or other symbols, but not white spaces)
- Example file: **example_events.csv** in folder **exercises/data_load**
 - Copy the file to the hard drive of your computer, e.g., to **My Docs** (drive **u**)
- Start V-Analytics (StudentUI)
- Menu “File” > “Load data”
- A dialog appears as on the right > Press button “ASCII file”
- In the following dialog, press “Browse”, then find and select the file **example_events.csv**.





This dialog appears after the file selection.

Reading data from ASCII format

Data sample:

```
MESSAGEID, LONGITUDE, LATITUDE, MESSAGEDATE, USERID, USERSCREENNAME, MESSAGE  
448970961187254273,-0.1330558955669403,51.52777099609375,27/03/2014 00:53:13,2180267  
448971342923452416,0.0024423201102763414,51.468544006347656,27/03/2014 00:54:44,463  
448973322433617920,-0.1328098624944687,51.55046081542969,27/03/2014 01:02:36,1987377  
448973731399237633,-0.02388888970017433,51.50583267211914,27/03/2014 01:04:13,984523  
448974176192585728,-0.04868140071630478,51.48075866699219,27/03/2014 01:05:59,407348  
448975589886935040,-0.1065528467297554,51.505592346191406,27/03/2014 01:11:36,199676  
448975595293396992,-0.09489571303129196,51.53363037109375,27/03/2014 01:11:38,410207  
448976049721077760,-0.2872936427593231,51.5728  
448977136834334720,-0.12102200090885162,51.514
```

Value separator: (N - numeric, C - character, L - logical, D - date)

Take field types from the line N

Take field names from the line N

Identifiers are in field (enter the name or the number of the field)

If there is no field with identifiers, the system will produce default identifiers from record numbers

Multiple lines may describe a single object

For instance, these may be points of a line or trajectory.
In such a case, the identifiers in the rows describing one and the same object must coincide.

Names of entities are in field (enter name or number or leave empty)

Coordinates of entities are in fields

X: Y: (enter names)

Circle radii are in field: (enter name)

OK Cancel

Check if the separator has been correctly recognised



Reading data from ASCII format

Data sample:

```
MESSAGEID, LONGITUDE, LATITUDE, MESSAGEDATE, USERID, USERSCREENNAME, MESSAGETEXT
448970961187254273,-0.1330558955669403,51.52777099609375,27/03/2014 00:53:13,2180267
448971342923452416,0.0024423201102763414,51.468544006347656,27/03/2014 00:54:44,463
448973322433617920,-0.1328098624944687,51.55046081542969,27/03/2014 01:02:36,1987377
448973731399237633,-0.02388888970017433,51.50583267211914,27/03/2014 01:04:13,984523
448974176192585728,-0.04868140071630478,51.48075866699219,27/03/2014 01:05:59,407348
448975589886935040,-0.1065528467297554,51.505592346191406,27/03/2014 01:11:36,199676
448975595293396992,-0.09489571303129196,51.53363037109375,27/03/2014 01:11:38,410207
448976049721077760,-0.2872936427593231,51.57285690307617,27/03/2014 01:13:26,2849688
448977136834334720,-0.12102200090885162,51.51415252685547,27/03/2014 01:17:45,144722
```

Value separator: ,

Take field types from the line N (N - number)

Take field names from the line 1

Identifiers are in field 1

If there is no field with identifiers, the system will produce default identifiers.

Multiple lines may describe a single object. In this instance, these may be points of a line or trajectory. In such a case, the identifiers in the rows describing one and the same object must be the same.

Names of entities are in field (enter name)

Coordinates of entities are in fields

X: LONGITUDE Y: LATITUDE (enter name)

Circle radii are in field: radius (enter name)

OK Cancel

Check if your data include coordinates

If your data do not include identifiers, leave this field empty (default identifiers will be generated).

Fill in this field if your data include names; not in our case.

Put the right column names here; the default names "X" and "Y" may not be right for your data. In our example, the right names are "LONGITUDE" and "LATITUDE".



Check the resulting specification.

Reading data from ASCII format

Data sample:

```
MESSAGEID, LONGITUDE, LATITUDE, MESSAGE DATE, USERID, USER SCREEN NAME, MESSAGE TIME  
448970961187254273,-0.1330558955669403,51.52777099609375,27/03/2014 00:53:13,2180267  
448971342923452416,0.0024423201102763414,51.468544006347656,27/03/2014 00:54:44,463  
448973322433617920,-0.1328098624944687,51.55046081542969,27/03/2014 01:02:36,1987377  
448973731399237633,-0.02388888970017433,51.50583267211914,27/03/2014 01:04:13,984523  
448974176192585728,-0.04868140071630478,51.48075866699219,27/03/2014 01:05:59,407348  
448975589886935040,-0.1065528467297554,51.505592346191406,27/03/2014 01:11:36,199676  
448975595293396992,-0.09489571303129196,51.53363037109375,27/03/2014 01:11:38,410207  
448976049721077760,-0.2872936427593231,51.57285690307617,27/03/2014 01:13:26,2849688  
448977136834334720,-0.12102200090885162,51.51415252685547,27/03/2014 01:17:45,144722
```

Value separator: (N - numeric, C - character, L - logical, D - date)

Take field types from the line N

Take field names from the line (N - numeric, C - character, L - logical, D - date)

Identifiers are in field (enter the name or the number of the field)

If there is no field with identifiers, the system will produce default identifiers from record numbers

Multiple lines may describe a single object

For instance, these may be points of a line or trajectory.
In such a case, the identifiers in the rows describing one and the same object must coincide.

Names of entities are in field (enter name or number or leave empty)

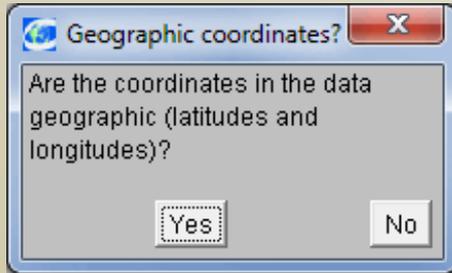
Coordinates of entities are in fields

X: Y: (enter names)

Circle radii are in field: (enter name)

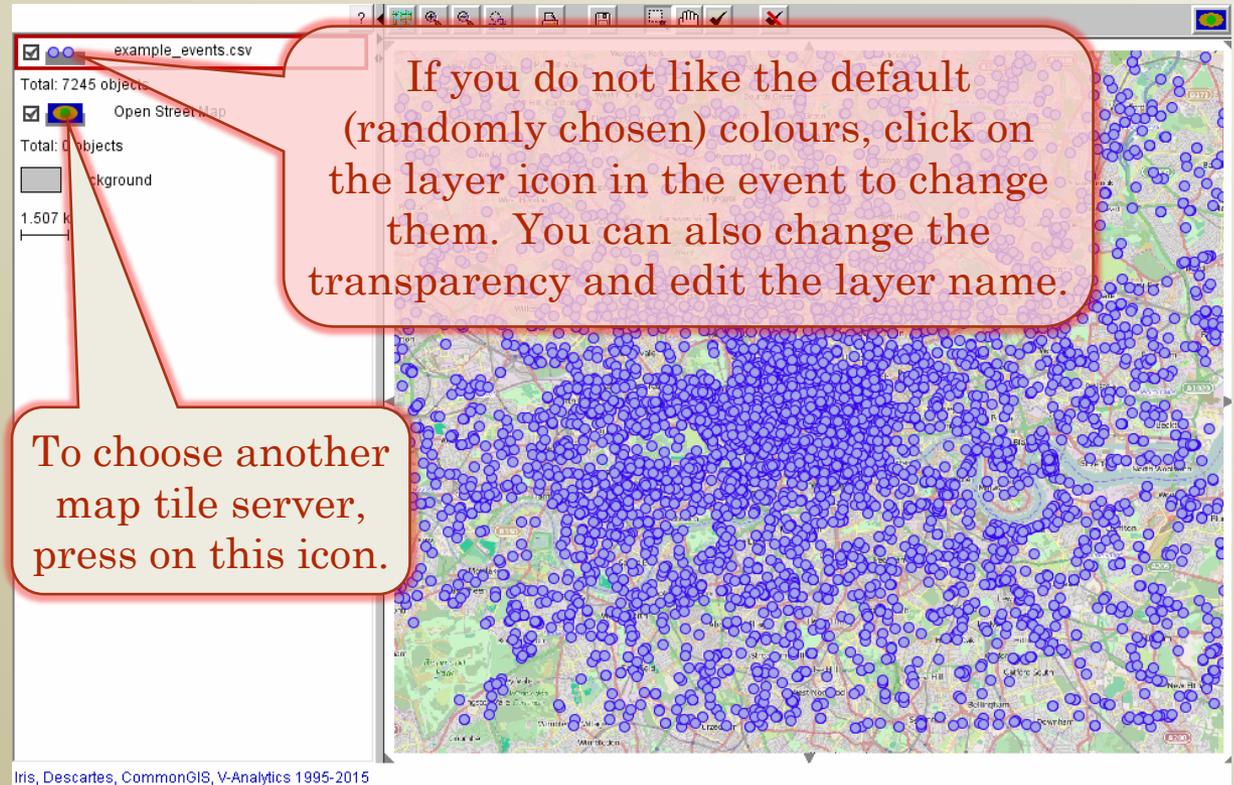
OK Cancel

Press after making and checking all settings



If the data that you load include coordinates, and no other spatial data have been loaded yet (the system window is empty), the system asks whether the coordinates are geographic. Reply correctly! In our example, we press “Yes”.

If the data contained coordinates, the system creates a map layer with point objects and puts this layer on the map (if the map did not exist, it is created). If the coordinates are geographic, the system automatically adds a map layer with Open Street Map tiles.



The system also creates a table with all data that have been loaded. The table can be seen through menu “Display” > “Table view”. Such a table is created also when the data do not contain coordinates.



2) Indexing of a table with time references

- The system does not yet know that the objects constructed from the data are spatial events. To let the system know this, we need to do *table indexing*:
 - Menu “File” > “Index a table with parameters or time references”

Select the column with the time references and press the “Select” button.

Table indexing; stage 1 of 3
Finding and extracting time references specified in one or more columns

Task Explanations Examples

Does the table contain one or more columns that specify the time moments the data in rows refer to?

Columns	Values
X	27/03/2014 00:53:13
Y	27/03/2014 00:54:44
LONGITUDE	27/03/2014 01:02:36
LATITUDE	27/03/2014 01:04:13
MESSAGEDATE	27/03/2014 01:05:59
USERID	27/03/2014 01:11:36
USERSCREENNAME	27/03/2014 01:11:38
MESSAGETEXT	27/03/2014 01:11:38
LOCATION	27/03/2014 01:11:38

Sort alphabetically

Time references are in columns:

Column name:	Format:	Meaning or template:
MESSAGEDATE	Simple value	year

Resulting column or parameter name: _____

This is a parameter Keep original column

Protract known values forward in time

(where values are missing, insert the values from previous time moments)

Template for displaying resulting dates/times: _____

OK

Press this button to see the values contained in the highlighted column. The values are shown in the list on the right.

These boxes must be left unchecked.

You need to specify the format of the time references. The next slide shows an example for dates and times.



We first replace “Simple value” by “Compound value”:

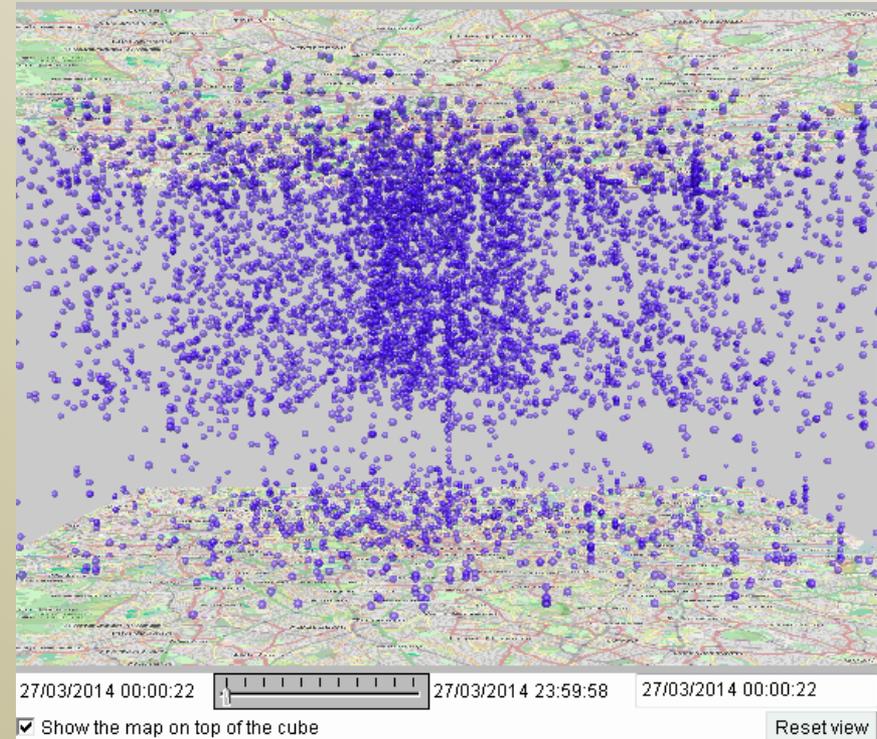
Column name:	Format:	Meaning or template:
MESSAGEDATE	Compound value	<enter the template>

Then we enter the date and time template:

Column name:	Format:	Meaning or template:
MESSAGEDATE	Compound value	dd/mm/yyyy hh:tt:ss

Note that minutes are represented by character ‘t’ since ‘m’ is used for months.

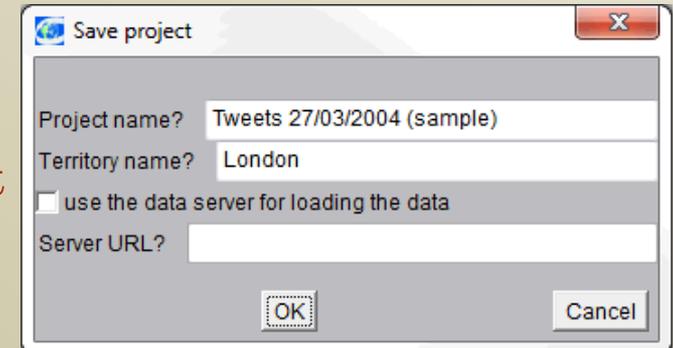
- After pressing OK, another dialog appears. We press “Cancel”.
- The system attaches time references to the point objects.
- To see the result of the indexing, we can visualise the events in a space-time cube.





3) Saving a project

- It took some time and effort for us to instruct the system how to treat the data. To avoid doing the same next time when we want to analyse the same data, we ask the system to store the instructions.
- Instructions are stored in special text files with extension .app, called project files or application files.
- To create a project file, go to menu File > Save project
- In a dialog that appears, enter a name for the project and a name for the territory, press OK.
- In the following file saving dialog, give a name to the project file (**pay attention that the extension must be .app**). The system saves the project.
- To check if that was done successfully, start another copy of the system and load the project in it.





Example of generating and
exporting new data



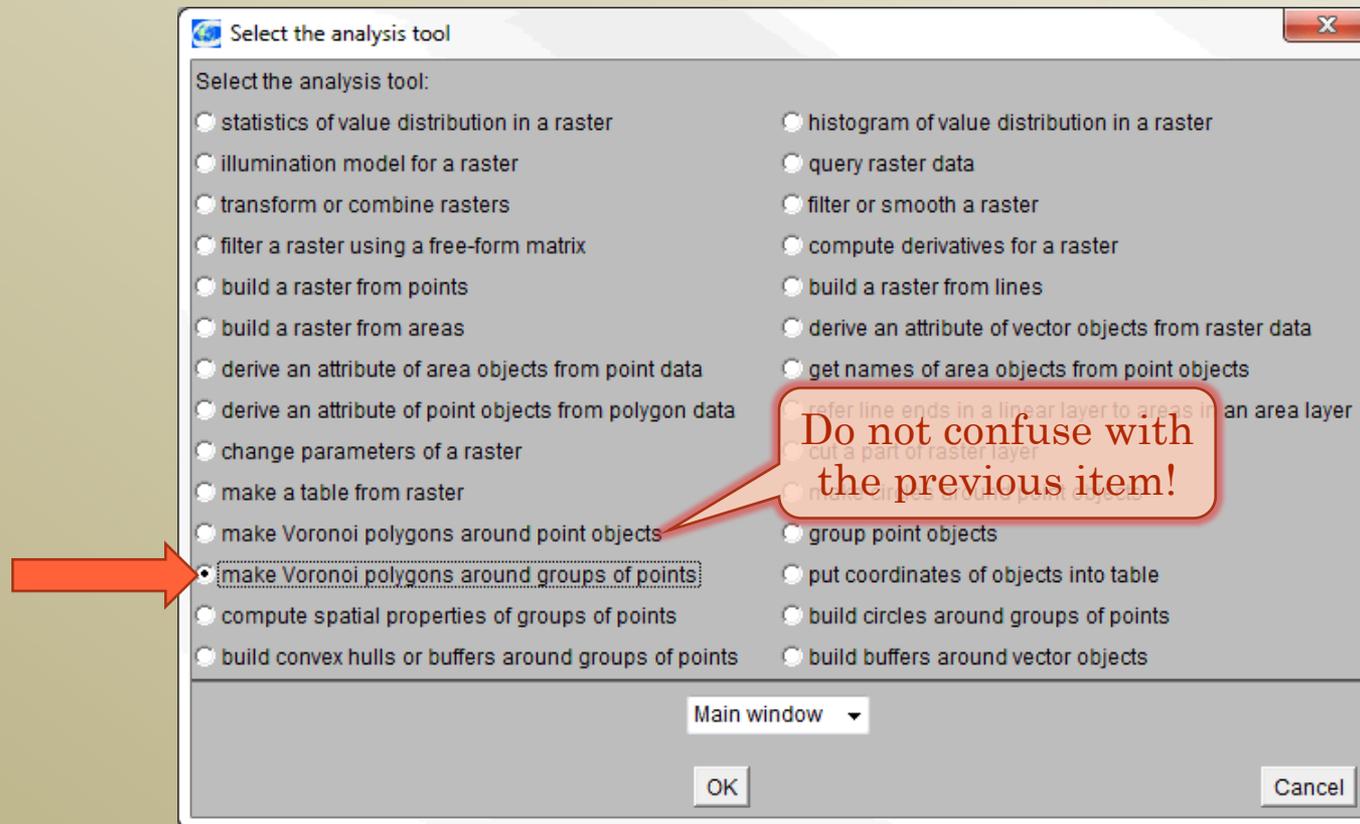
Exporting data

- In the process of analysis, you may generate new data, such as
 - New attributes (columns) in a previously existing table
 - A new table
 - A new map layer
- The new data reside in the system's internal structures in the RAM. They will be lost when you quit the system or load another project.
- If you may need the generated data later, or you want to apply other software to these data, you need to export the data from the system.
- The exporting functions are accessed through File > Export data.
- Important note: Saving a project does not automatically export data! You need to export valuable secondary data **explicitly**!
- After exporting, you can save the project. The system will write information about the exported data into the project file.



Generating new data: territory tessellation

- We use the previously loaded project with the spatial events.
- Menu “Calculate” > “Spatial calculations”
- Select the function “Make Voronoi polygons around groups of points”





After selecting the map layer with the point objects, you get the following dialog:

The 'Group points' dialog box shows the following settings:

- Maximum group radius?: 1000
- Refine the division by subdividing unbalanced (*) groups
* where the group mass centre is not at the geometric centre
Minimum group radius: 100
- Refine the division by subdividing large groups
Minimum number of points in a group to be divided: 100
Minimum group radius: 100
- X-extent: 26563
- Y-extent: 21914
- Scale: 1511.10 m
- make a map layer with the centroids of the groups
- introduce additional generating points for Voronoi cells in empty areas

Buttons: OK, Cancel

Make the following settings:

The 'Group points' dialog box shows the following modified settings:

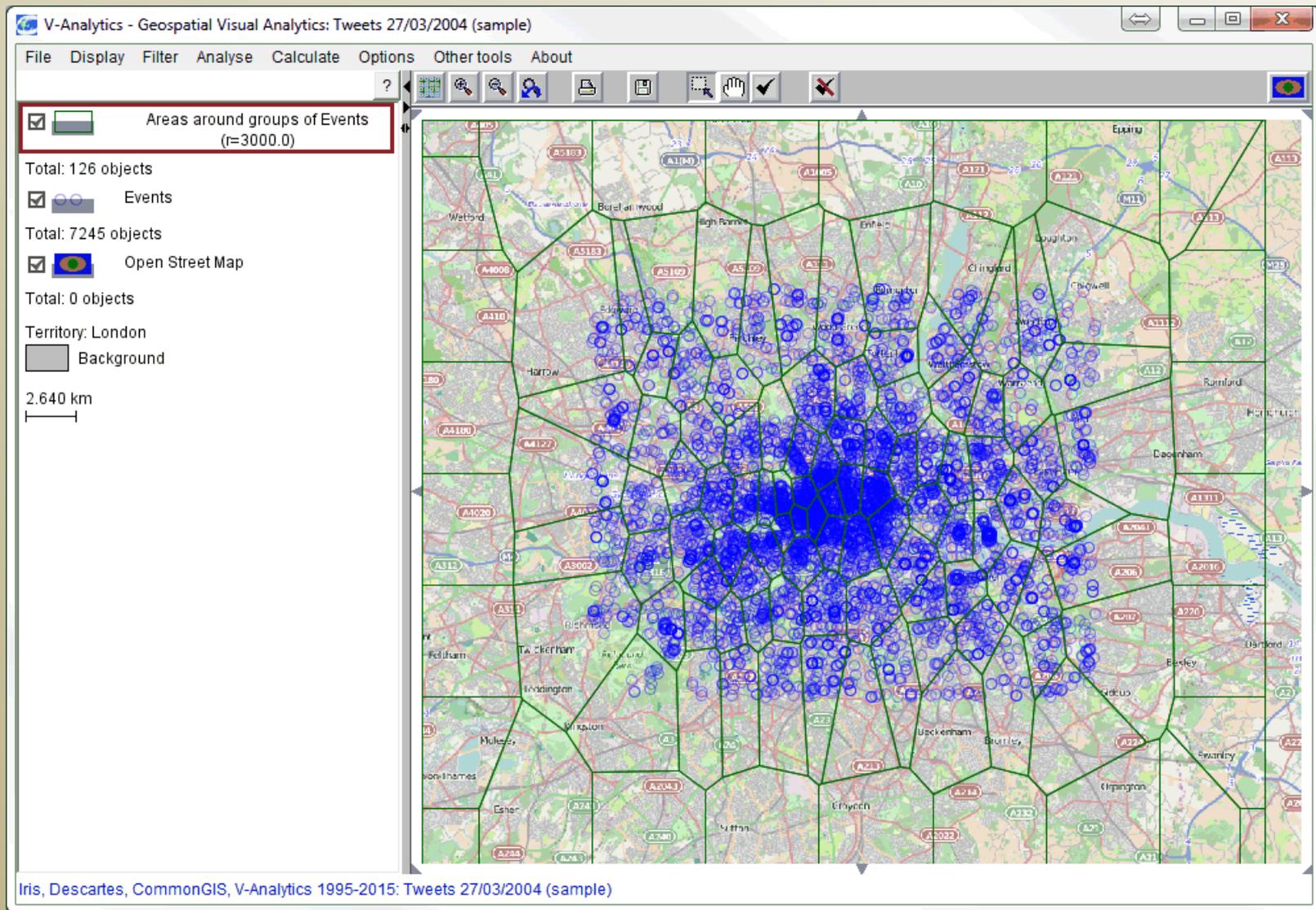
- Maximum group radius?: 3000
- Refine the division by subdividing unbalanced (*) groups
* where the group mass centre is not at the geometric centre
Minimum group radius: 1500
- Refine the division by subdividing large groups
Minimum number of points in a group to be divided: 500
Minimum group radius: 1000
- X-extent: 26563
- Y-extent: 21914
- Scale: 1511.10 m
- make a map layer with the centroids of the groups
- introduce additional generating points for Voronoi cells in empty areas

Buttons: OK, Cancel

If you need to use this tool in your analysis, you may need to try several combinations of settings and choose the most suitable result.

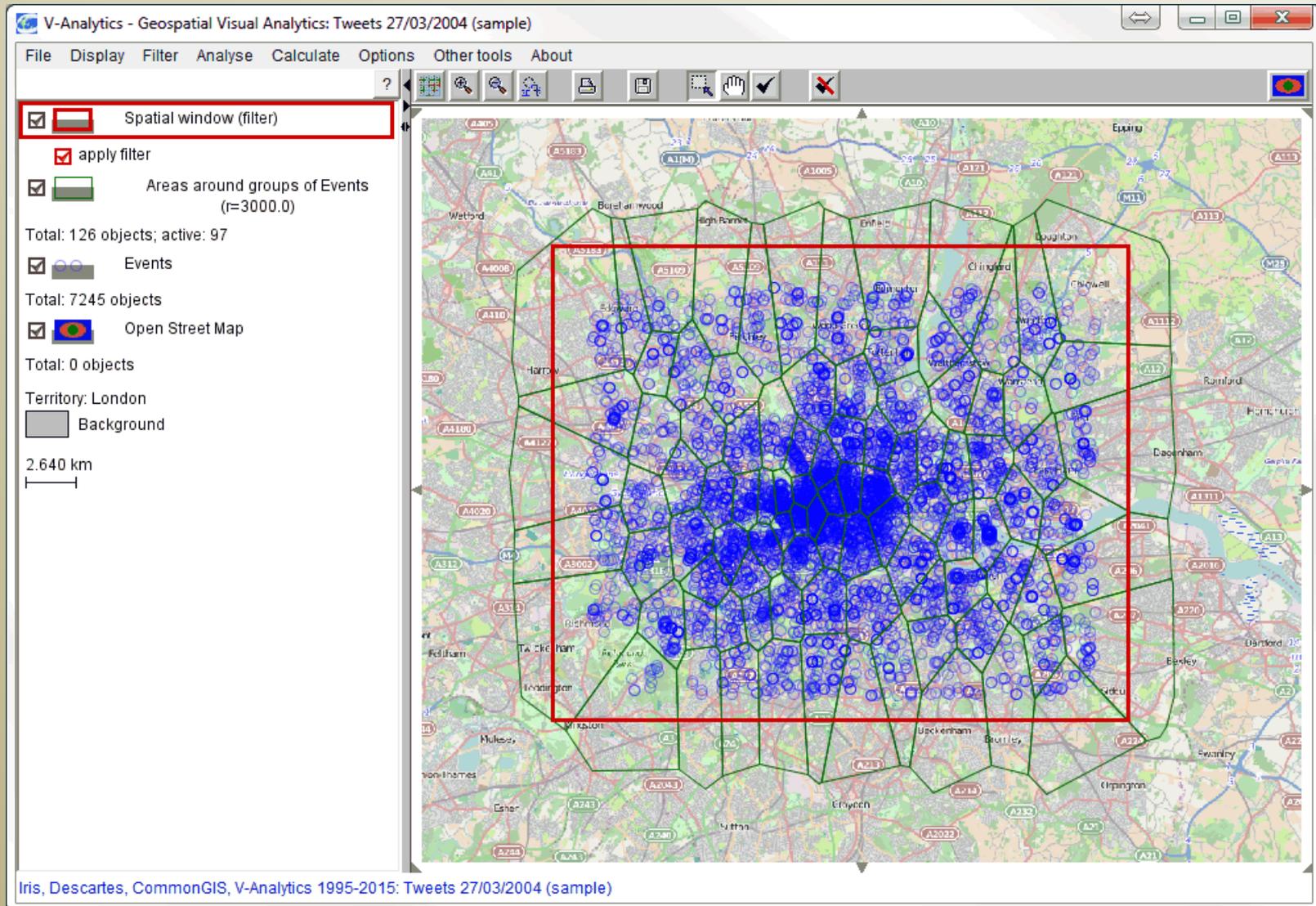


The system generates a new map layers with polygons:





Before exporting the new data, you may wish to filter them, e.g., by a spatial window. **Export functions are always applied to active data** (i.e., those that are not filtered out).





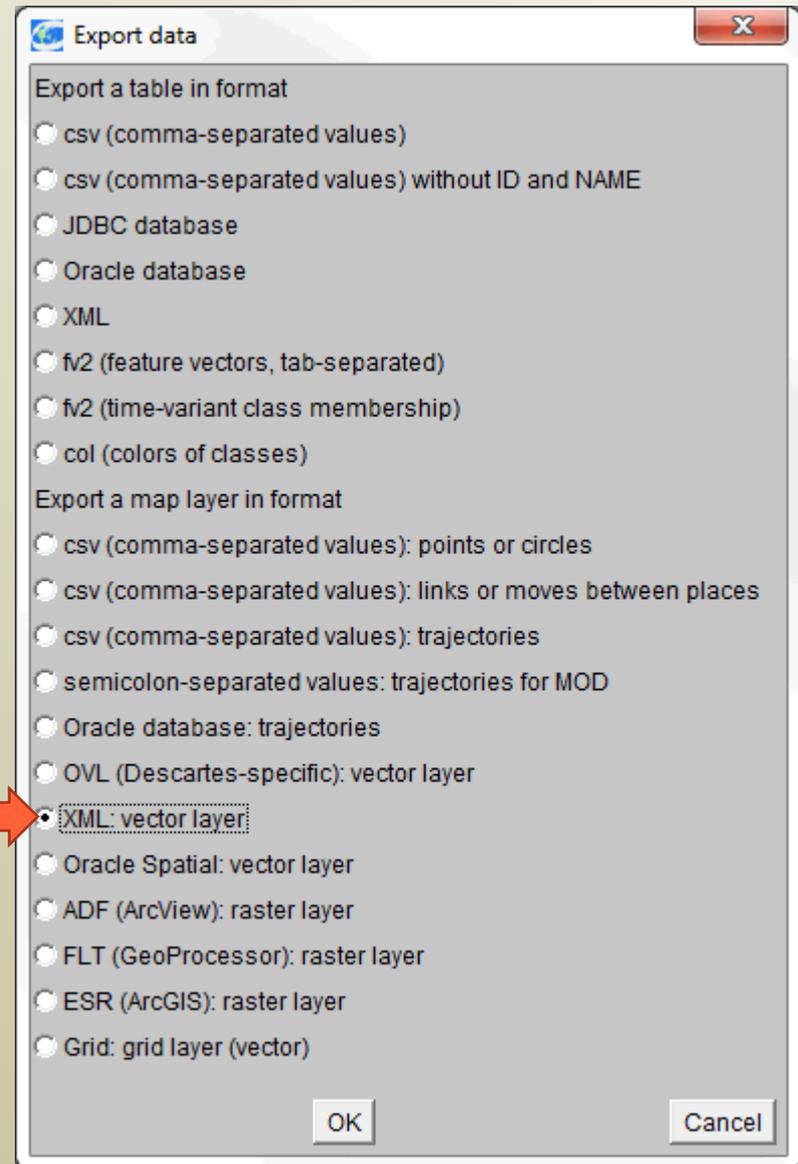
Exporting the polygons

Menu “File” > “Export data”

In this dialog, you need to find a suitable format for the data you wish to export. The easiest case is a table or a layer with point objects – these can be stored in a text file in CSV (comma-separated values) format. Data can also be exported to a database.

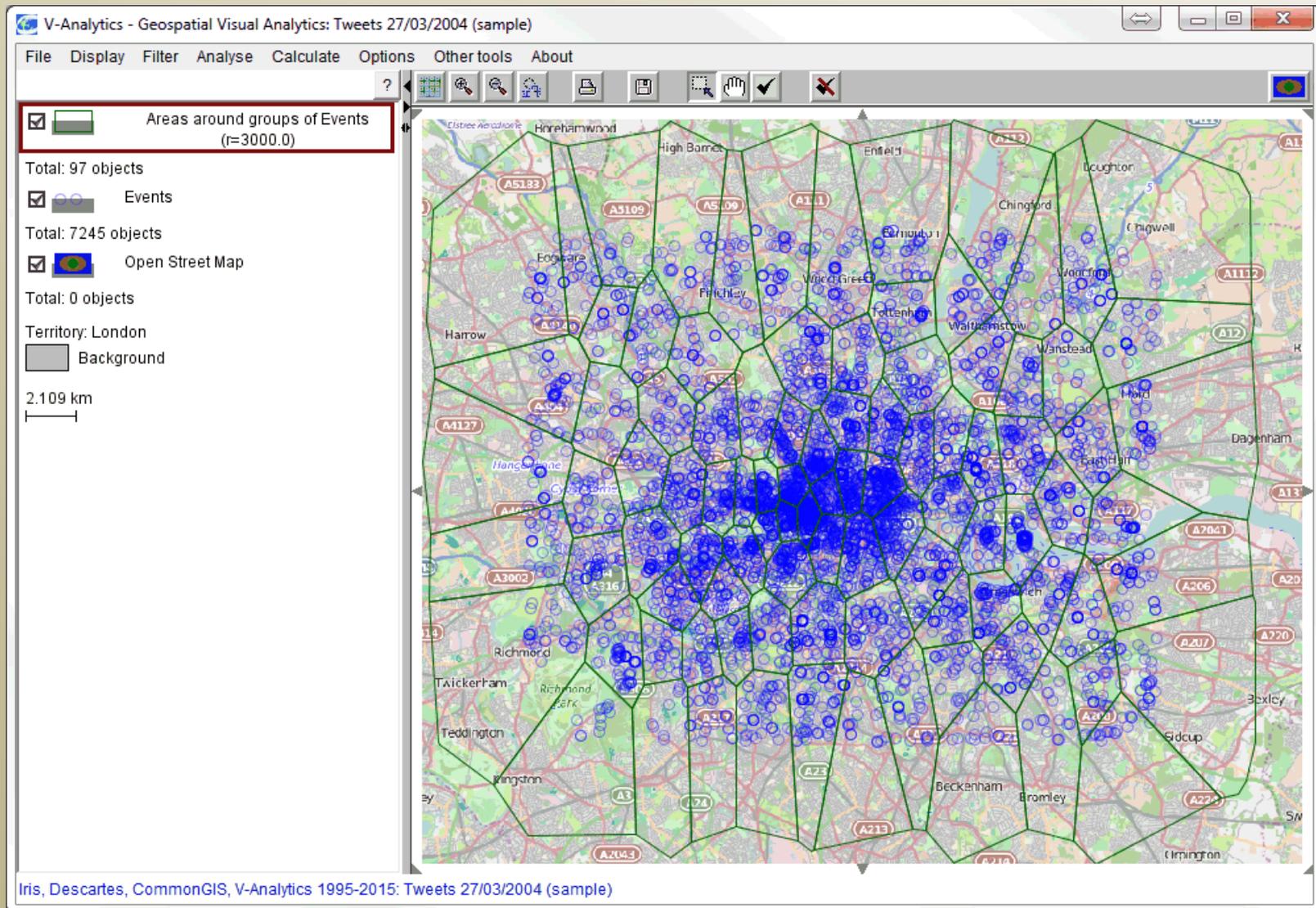
A map layer consisting of lines or polygons can be stored in XML format. This is **not** a standard format used also in other software; it is readable only by the V-Analytics system. However, the format is easy to understand and to parse. It is not very difficult to write a parser that will transform this format to something else, if you need this.

After exporting, save the project.





Next time, when we load the project in the system, we get a map like this:



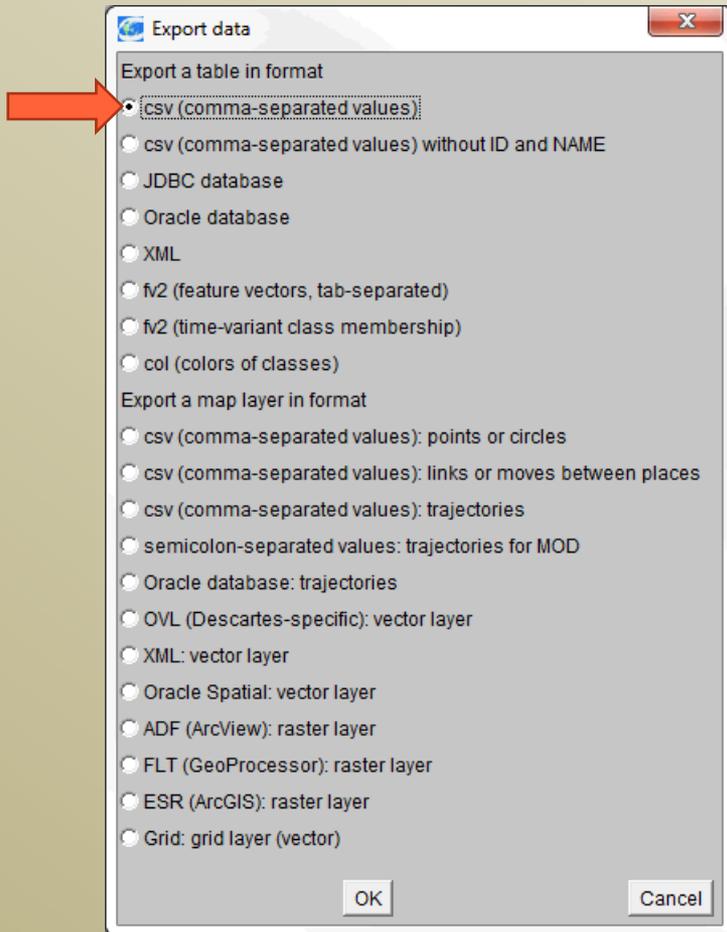


Exporting to XML format: some hints

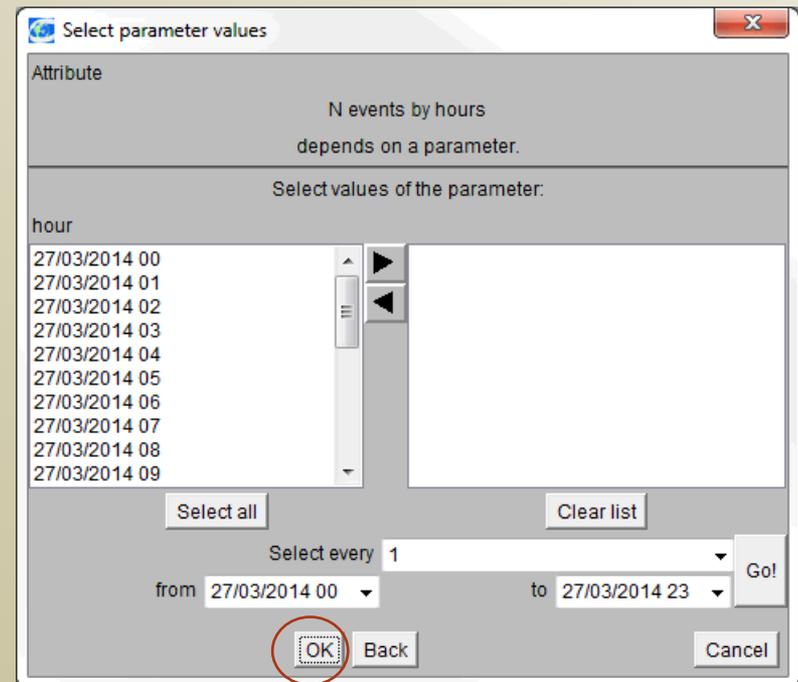
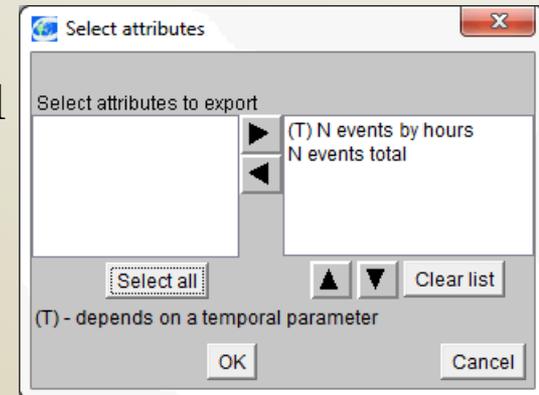
- The XML format takes much space on a hard drive and also much time for loading.
- When you have generated some (especially many) attributes associated with your polygons, e.g., time series of event counts, it is recommended to store the attributes separately from the polygons.
 - Polygons > XML file
 - Attributes > CSV file
 - When you save the project, the system will store information about the association between the layer with the polygons and the table with the attributes.
- When you start exporting polygons that have thematic attributes, the system asks what attributes you want to export. Do not select any attributes. The resulting XML file will only contain the spatial data.
- Next, export the table with the attributes.

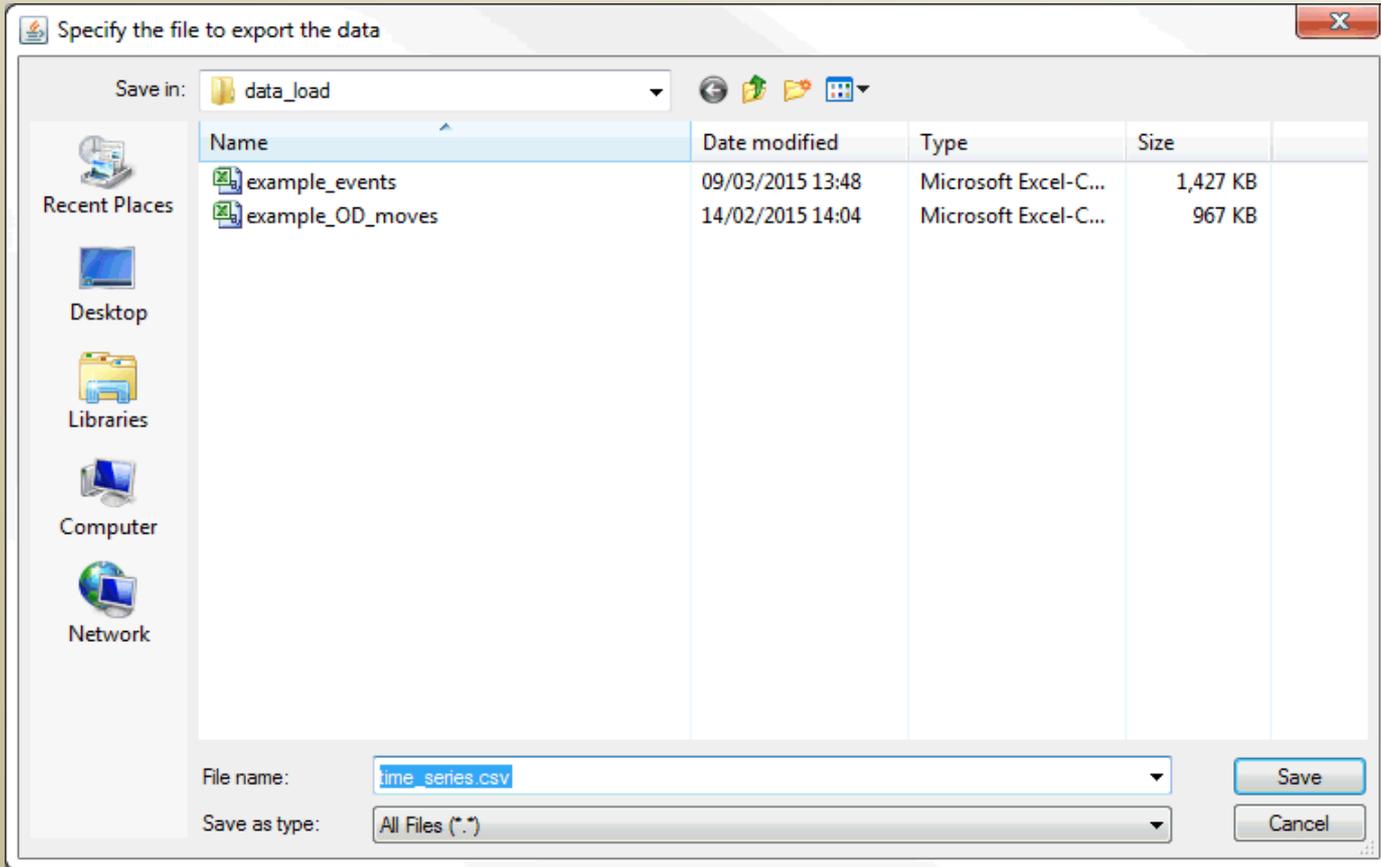


Exporting a table with thematic attributes



The CSV format is understood by Excel and many other software systems.





After exporting the data, save the project.

Next time, when you open the project, the system will load both the polygons and the table with their attributes and properly link each polygon with the corresponding row of the table.



Example of loading OD moves



Data format

- OD movement data can be loaded from a text (ASCII) file with separators (in particular, CSV – comma-separated values format).
- From the system's viewpoint, an OD move is a trajectory consisting of 2 points. Hence, the data format for OD moves is the same as for trajectories in general.
- In a text file, each trajectory is specified by a sequence of lines, each line describing one point. Hence, an OD move is specified by two lines.
 - The lines belonging to the same trajectory must be together in one block. There should be no intrusion of points from other trajectories.
- Each trajectory (OD move) must have a unique identifiers. All data lines describing a trajectory must include the trajectory identifier.
 - The identifiers allow the system to unite multiple points of one trajectory (or two points of one move) into a single spatio-temporal object.



Example of loading OD moves

- We start the system anew.
- Menu “File” > “Load data” > “ASCII file” > browse for the file > select example_OD_moves.csv
- Fill in the dialog fields as shown.

This box must be checked for trajectories or OD moves.

Reading data from ASCII format

Data sample:

trID	trN	pldx	X	Y	time	duration
14284178	1	1	-0.19475795	51.493343	25/07/2012 00:03:10	0
14284178	1	2	-0.1811909	51.490944	25/07/2012 00:23:31	0
14287769	2	1	-0.053177	51.515976	25/07/2012 18:19:24	0
14287769	2	2	-0.053177	51.515976	25/07/2012 18:23:10	0
14287775	3	1	-0.11676421	51.513622	25/07/2012 20:09:55	0
14287775	3	2	-0.10802847	51.518826	25/07/2012 20:15:25	0
14287780	4	1	-0.11282408	51.50379	25/07/2012 18:31:44	0
14287780	4	2	-0.09177378	51.504627	25/07/2012 18:59:16	0
14287784	5	1	-0.077121325	51.511597	25/07/2012 17:24:26	0

Value separator: ,

Take field types from the line N

Take field names from the line N 1

Identifiers are in field 1

If there is no field with identifiers, the system will produce default identifiers from record numbers

Multiple lines may describe a single object

For instance, these may be points of a line or trajectory.
In such a case, the identifiers in the rows describing one and the same object must coincide.

Names of entities are in field (enter name or number or leave empty)

Coordinates of entities are in fields

X: X Y: Y (enter names)

Circle radii are in field: radius

OK

The presence of identifiers is important!

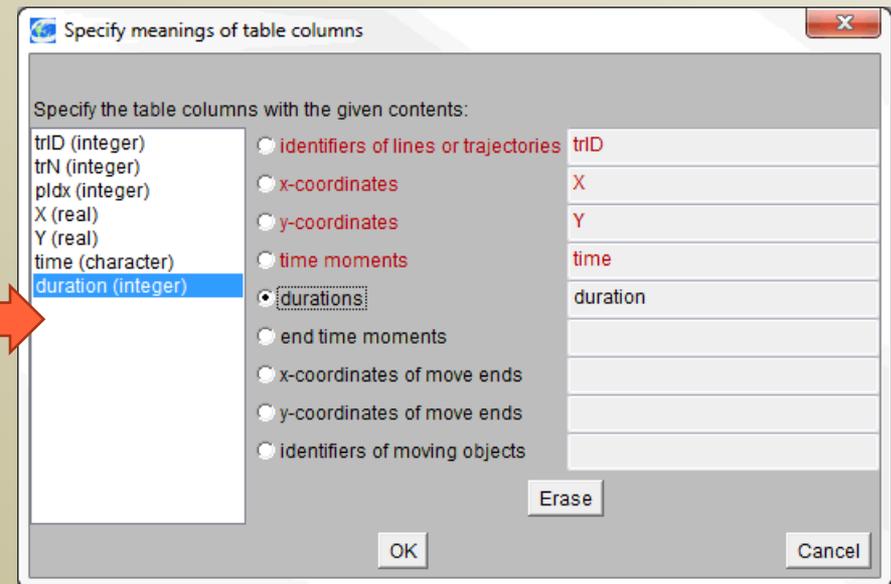
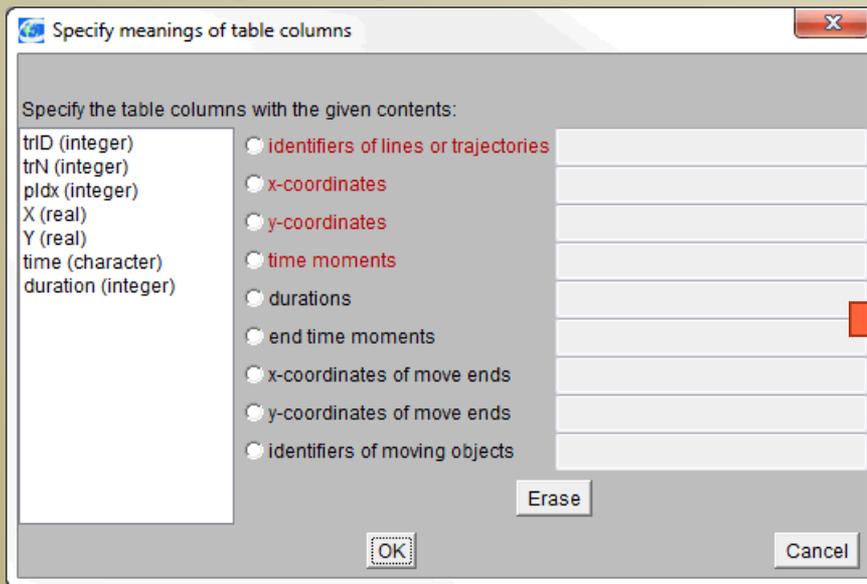
You do need to fill in these fields.

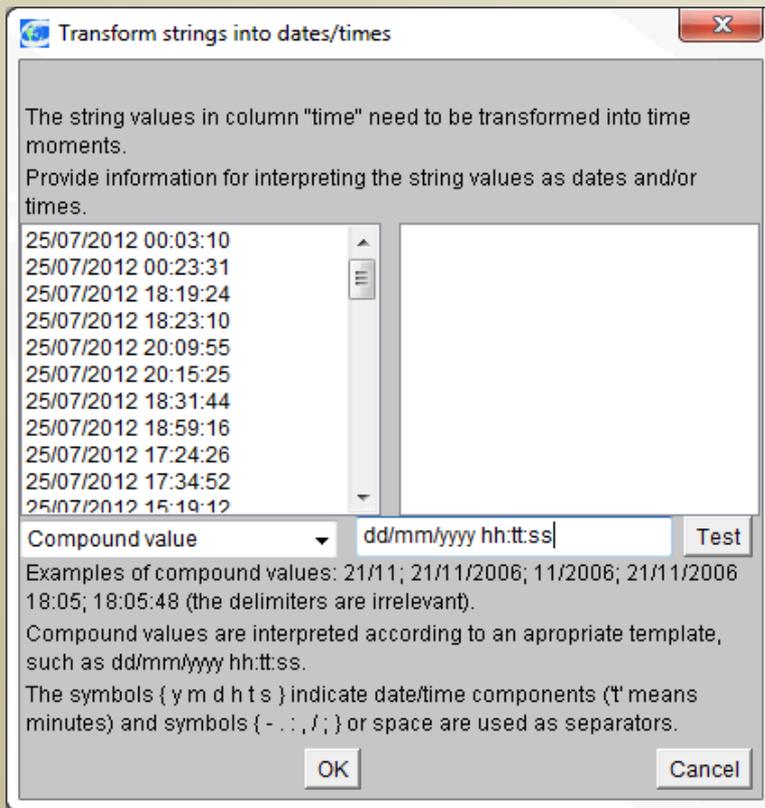
- After loading the data, no visible results appear. The data are put in a table.



Data indexing for OD moves or trajectories

- Menu “Other tools” > “Explore movement data”
- A dialog asks you to select a table with movement data. Select the table you have just loaded.
- A dialog appears in which you need to specify what table columns contain identifiers, coordinates, and time references.
 - Red: mandatory information, black: optional
 - Select a column name in the list, then select the radio button labelled by the corresponding meaning.

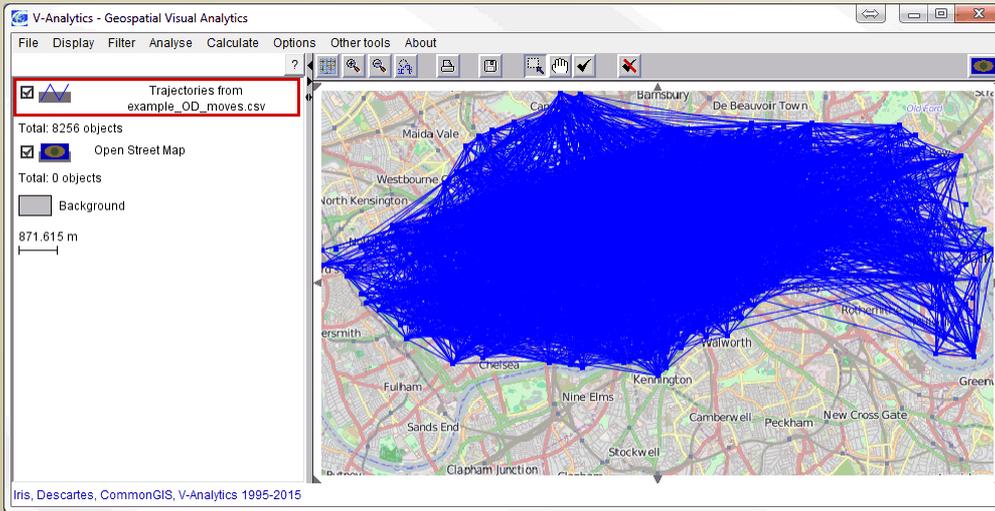




Next, the system asks how to transform the strings specifying the times into appropriate date and time values. The template for the transformation is specified in the same way as for events in example 1. Press OK after finishing.

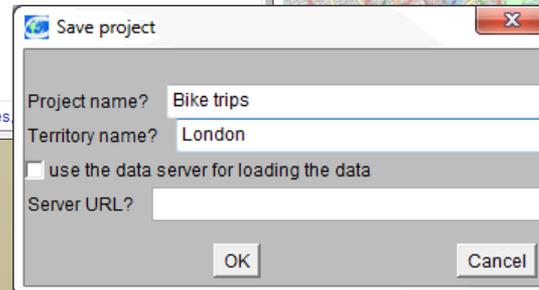
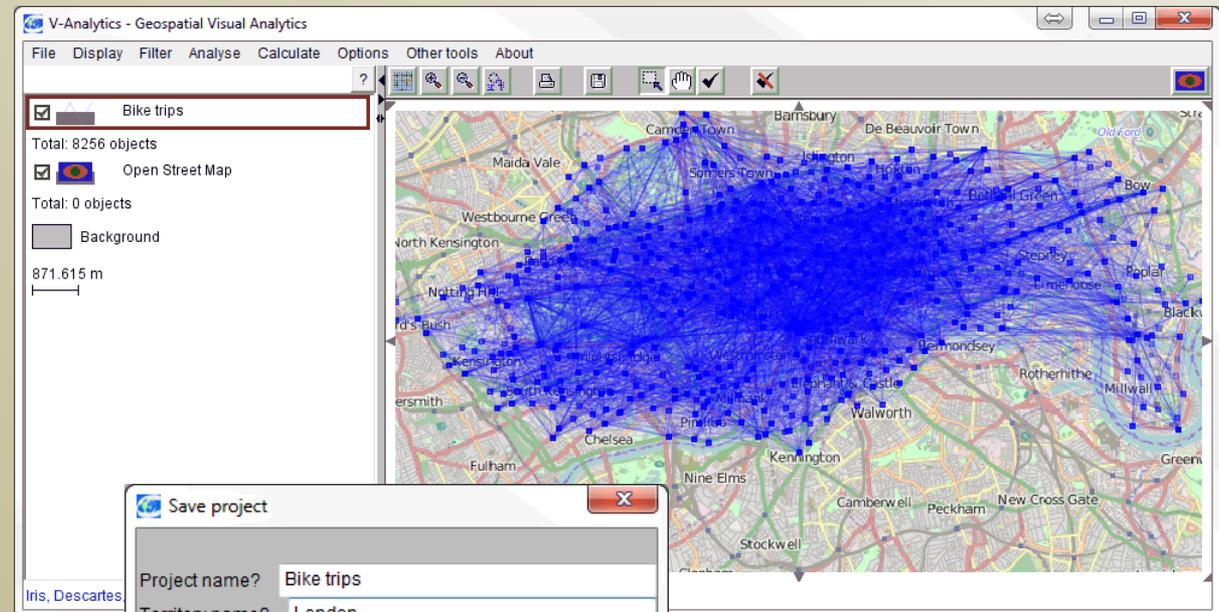
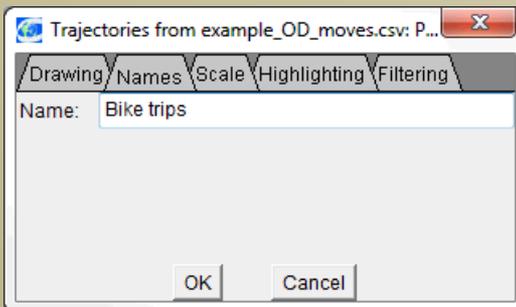
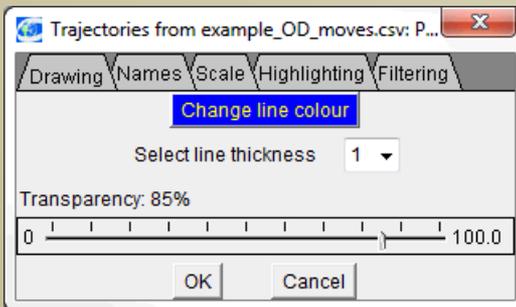
As in the example with events, the system will ask you whether the coordinates are geographic (longitudes and latitudes). For this example, press “Yes”.

The system creates a map with trajectories and displays a dialog “Explore movement data”. Press “Cancel” in the dialog.



You may change the layer colour, level of transparency, and/or name after clicking on the icon left to the layer's name in the map legend.

Now you can save the project, i.e., create a project description file analogously to the example with the events.





See also the explanatory texts
and videos

<http://geoanalytics.net/V-Analytics/>