Lecture 04

Space and Time

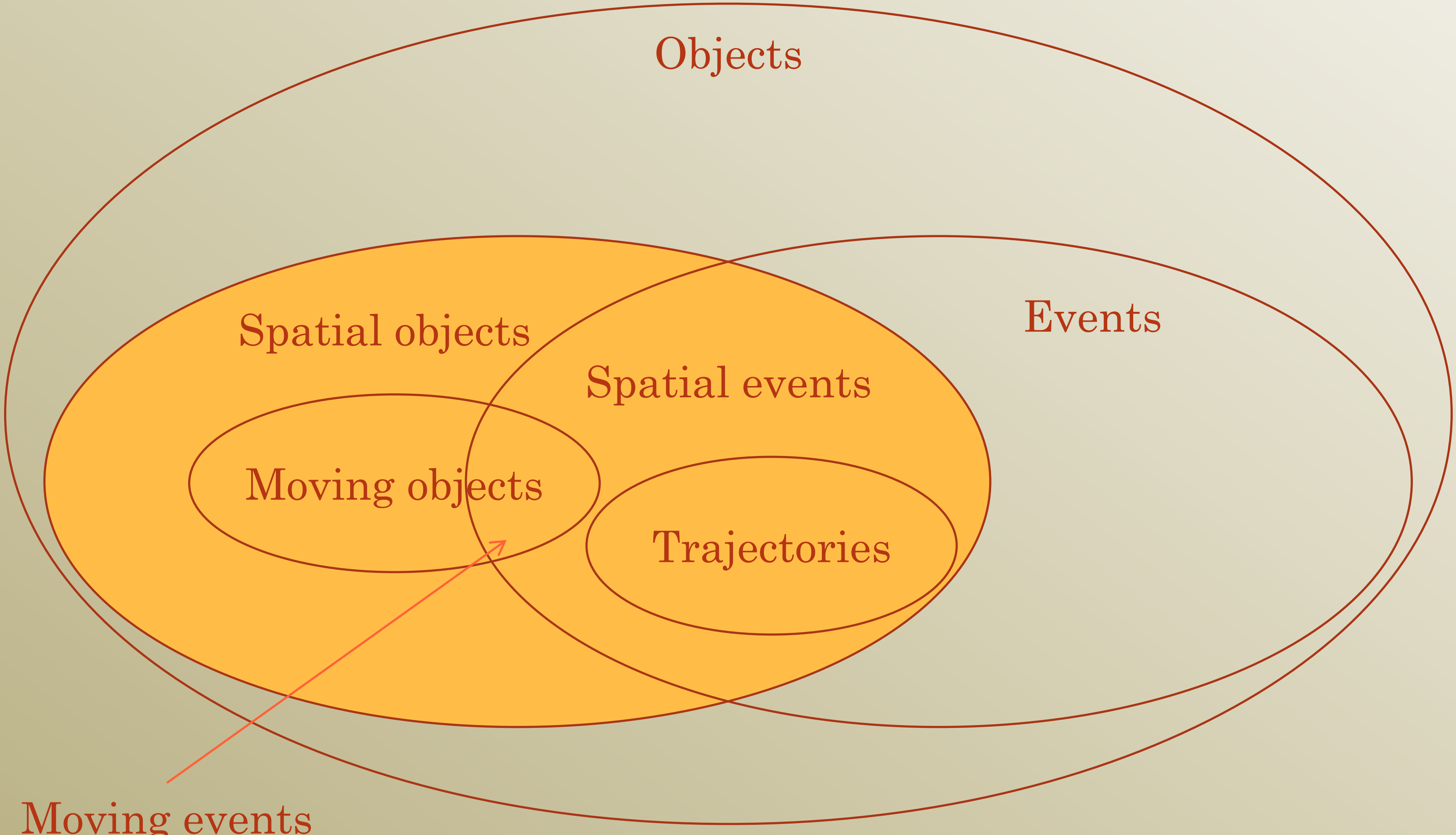
given by
prof. Gennady Andrienko and
prof. Natalia Andrienko

Content and objectives

- In addition to previously considered spatial events and spatial time series, we consider a new type of spatio-temporal data: OD moves.
- You will learn which of the earlier studied visualisation and analysis methods are applicable to OD moves and understand why specific approaches are additionally required for this data type.
- You will see further applications of density-based clustering with distance functions specific for OD moves and specific ways of spatio-temporal aggregation of OD moves.
- You will understand the relationships between different types of spatio-temporal data and the possibilities for transforming one type into another.

Classes of objects *(a reminder)*

Venn diagram



Types of objects

based on their properties and attributes

- **Generic objects**
    - Wine varieties, car models, plant specimens, …
- **Spatial objects**
  - Have locations in space ⇒ the attributes include the location
    - Districts, buildings, streets, rivers, …
- **Temporal objects**, a.k.a. ***events***
  - Have limited existence time ⇒ the attributes include the time of existence
    - Instant objects: have no duration; only the appearance time needs to be specified
      - Tweet postings, bank transactions, …
    - Durable objects: have duration; the attributes need to include the time of appearance + the time of disappearance or the duration
      - Holidays, electoral campaigns, classes, breaks, TV shows, …

Types of objects *(continued)*

spatio-temporal objects

- **Spatial events**
  - Events that have location in space ⇒ the attributes include the spatial location and the existence time (instance or interval)
    - Lightnings, geolocated tweet postings, earthquakes, traffic jams, …
- **Moving objects**
  - Spatial objects whose locations change over time ⇒ the attributes include spatial locations at different times, i.e., time series of spatial locations, called ***trajectories***
    - People, animals, vehicles, storms, oil spills, …
- **Trajectories** can be treated as spatio-temporal objects
  - Spatial location = spatial footprint
  - Existence time = time from the beginning till the end of the movement
  - Other properties → attributes: shape, travelled distance, mean speed

Classes of spatio-temporal data

- Spatial time series: time series referring to spatial objects or locations *(can also be considered as spatial objects)*
  - 2 referrers: space (set of spatial objects) × time (set of time steps)
  - One or more thematic attributes
- Spatial event data: contain attributes specifying spatial positions and existence times of spatial events.
- Movement data: contain attributes specifying spatial positions of *moving objects* at different times
  - Origin-destination (OD) data: specify only the positions and times of trip starts and ends; intermediate positions are not available.
  - Trajectories: specify also the intermediate positions of the moving objects at different times
    - Can be seen as time series of spatial positions

Origin-destination movement data (OD moves)

Introduction of a new data type

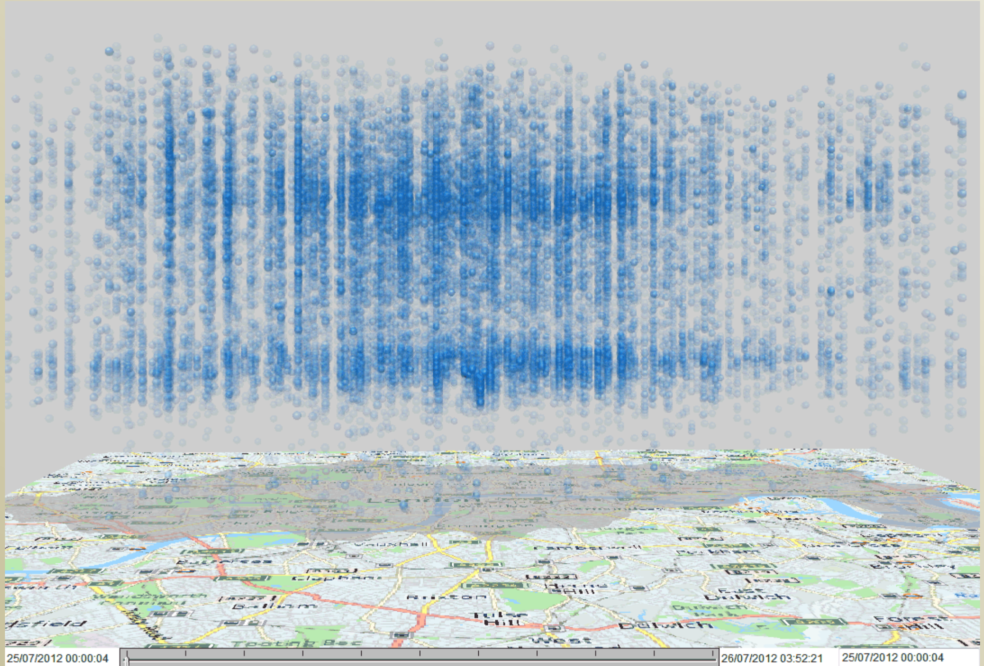
A running data example

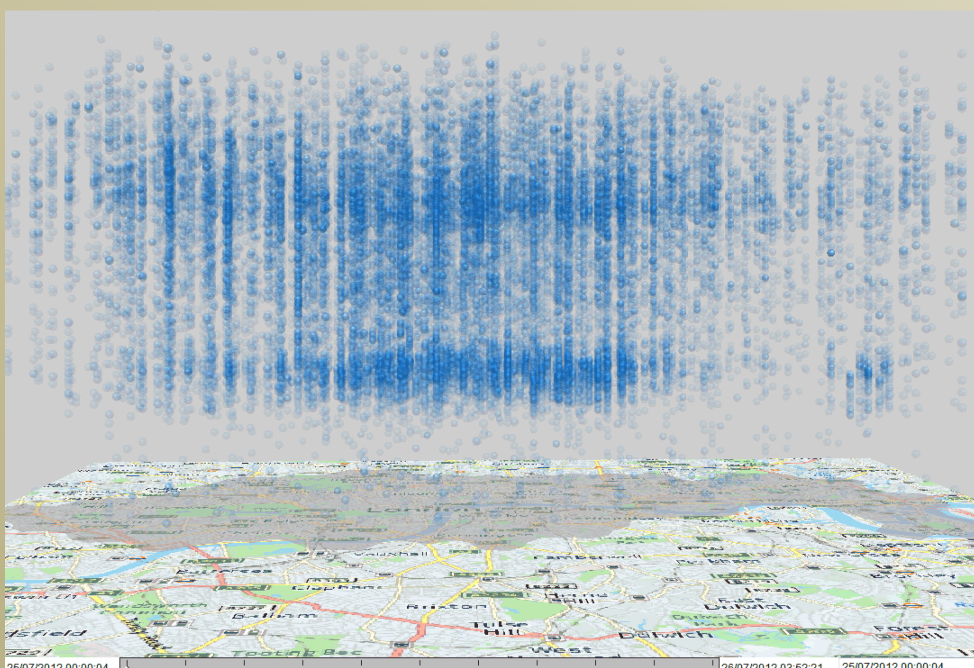
Use of shared bikes in London

- Barclays Cycle Hire
- 569 docking stations to pick up and return bicycles
- The data are publicly available from **Transport for London** (http://www.tfl.gov.uk/info-for/open-data-users/our-feeds)
- Data describe the journeys made with the rented bikes:
  - Journey ID
  - Bike ID
  - Start date & time
  - End date & time
  - Start docking station ID
  - End docking station ID
    - The geographic coordinates of the docking stations are known and can be joined with the transaction data

Relation of OD moves to spatial events

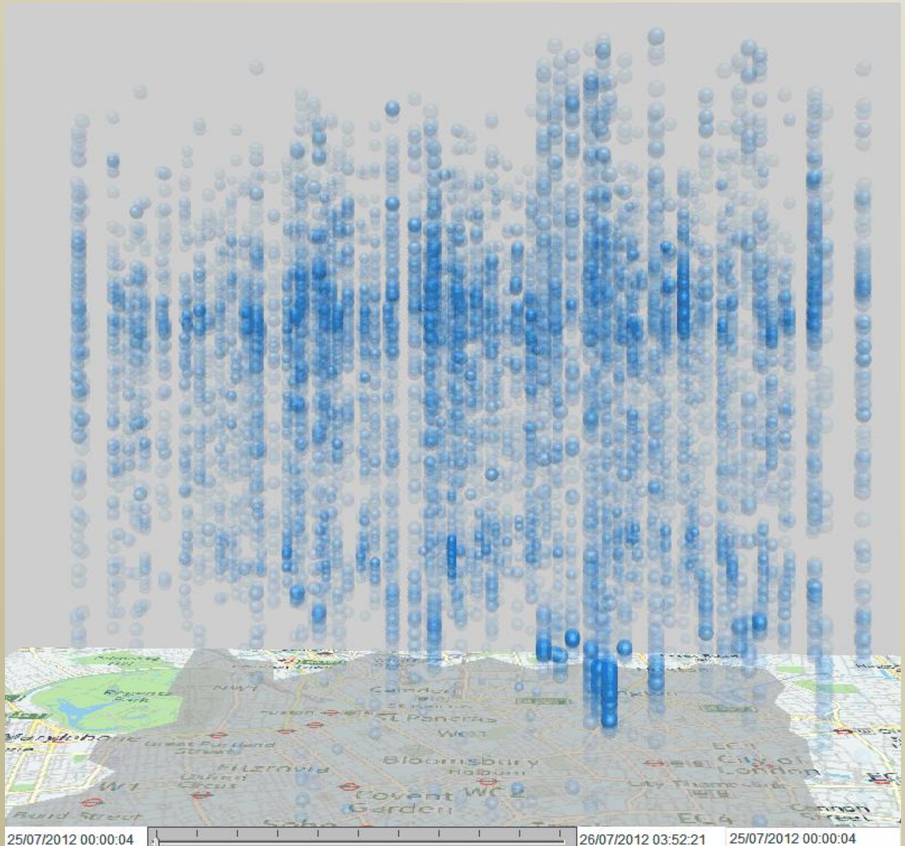
- Spatial events are objects having positions in space and times of occurrence or existence (= positions in time).
  - Space and time (and space × time) can be treated as containers of events.
  - Most types of synoptic tasks are concerned with the spatio-temporal distribution of events or values of their thematic attributes.
- An OD move includes 2 instant spatial events: start and end.
  - The spatio-temporal distributions of these events may be of interest in analysis.
  - The general types of synoptic tasks defined for spatial events in general apply to these events. The analysis methods suitable for spatial events can be used for the trip start and end events.

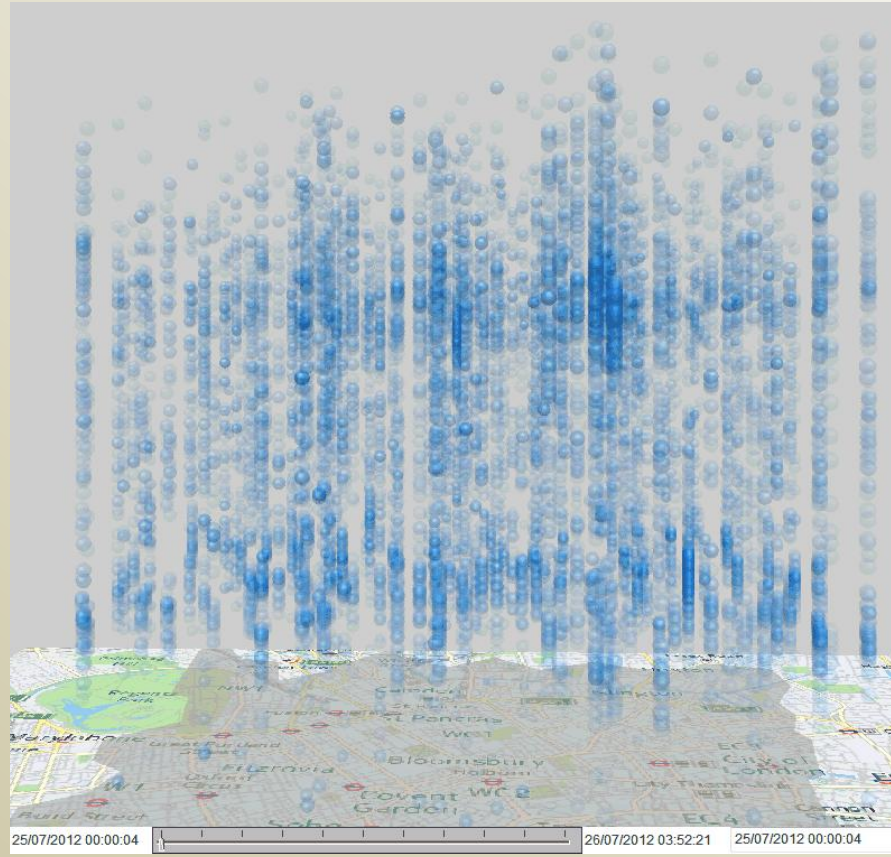




Example: instant point events extracted from the London bike trip data from Wednesday, 25/07/2012. The upper STC image shows the spatio-temporal distribution of the trip start events and the lower image represents the trip end events.

There are noticeable differences on the east; the other parts of the territory need to be explored with the help of spatial filtering, to decrease the display clutter and overplotting.





After applying spatial filtering and zooming in to the central part of the territory, the differences between the spatio-temporal distribution of the start (left) and end (right) events in this part become more vivid.

All visualisation and analysis methods applicable to point events can be used for analysis of the trip start and end events.

Relation of OD moves to spatial events

(continued)

- An OD move as a whole is a spatial event.
  - It has existence time = the time of the trip.
  - It has a position in space consisting of the start and end positions.
    - Such discontinuous spatial positions also occur for other spatial objects, e.g., a university.
    - The spatial position of a move can be represented by a directed line segment (vector), which is a spatial object.
- However, the visualisation and analysis methods designed for spatial events may not be applicable to this kind of events.
  - Most methods assume that the spatial positions of the events do not change during the time of event existence.
  - Moreover, most of them assume that the spatial positions of the events can be represented as points in space.
- Hence, OD moves require special approaches.

Synoptic tasks for OD moves

addressing the spatio-temporal distribution

- Describe the spatio-temporal distribution of the set of moves and their thematic attributes.
  - Thematic attributes: duration, spatial direction, displacement distance, transportation mode, trip purpose, characteristics of the moving object, …
- Find occurrences of particular distribution patterns in space × time.
  - E.g., spatio-temporal concentrations of move starts or move ends, spatio-temporal concentrations of similar moves (with close starts and ends), spatio-temporal co-occurrence of opposite moves, …
- Compare the spatio-temporal distributions of OD moves and their thematic attributes in different time periods, different parts of the space, …
- Relate the spatio-temporal distribution of OD moves and their thematic characteristics to behaviours of other spatial, temporal, and spatio-temporal phenomena.

Synoptic tasks for OD moves

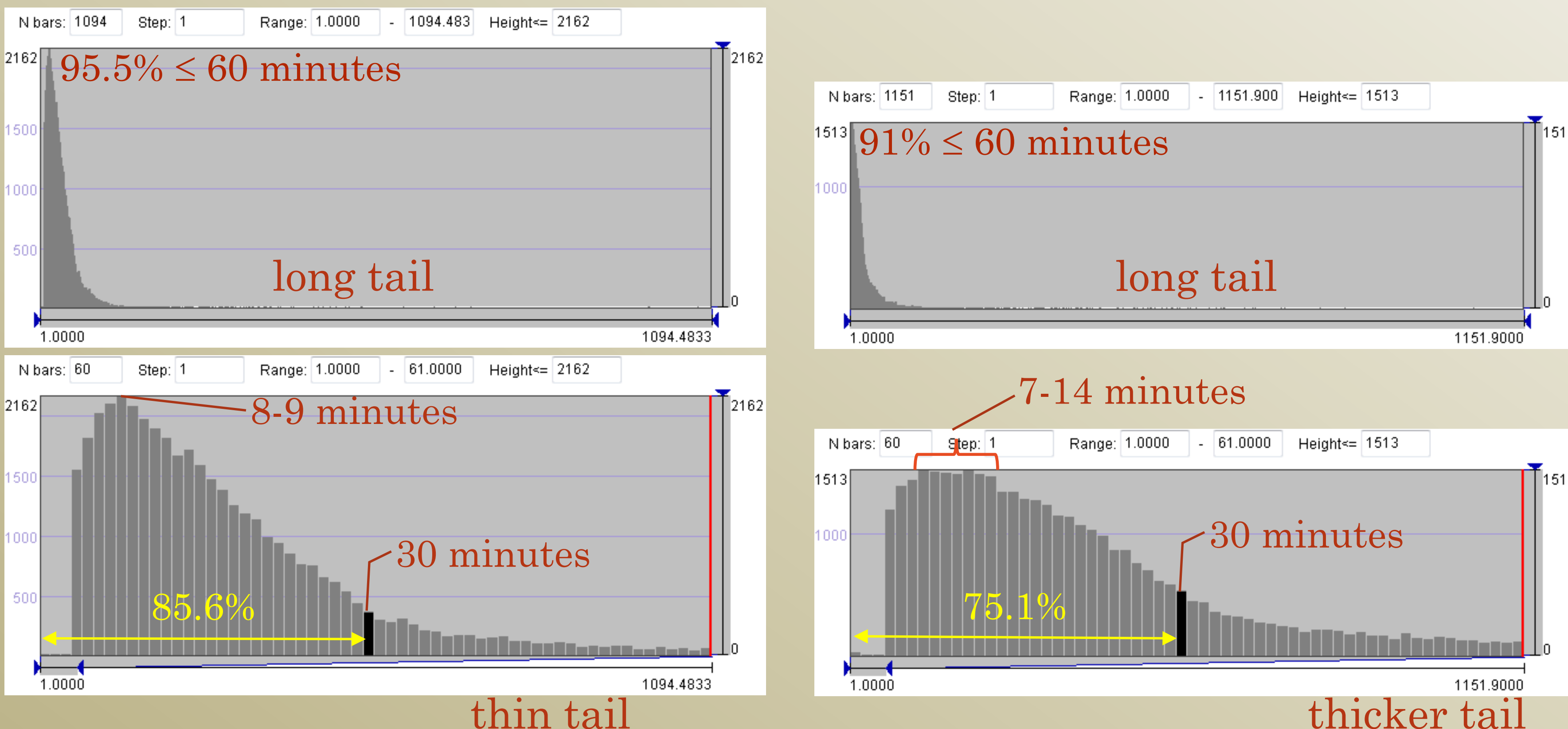
addressing the distribution of the thematic attributes over the set of moves

- Describe the distribution of values of the thematic attributes over the set of moves (treated as generic objects)
  - e.g., frequency distribution of the values and value combinations
- Find subsets of moves with particular combinations of thematic characteristics
  - Particularly, find where these subsets are in space and time.
- Compare the distributions of thematic characteristics for different subsets of moves
  - Particularly, moves that occurred in different time periods or in different parts of the space.
- Relate the distributions
  - of different thematic attributes
  - of thematic attributes and the spatial and temporal positions

OD moves as generic objects

Examples of analysing the attribute value distributions

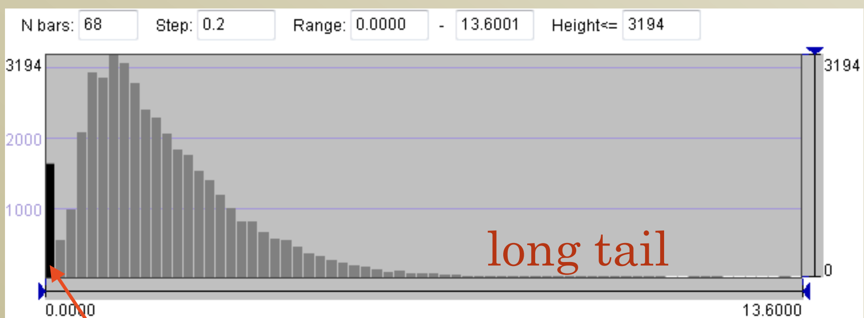
Frequency distributions of the bike trip **durations** on 25/07/2012 (Wednesday; 41,380 trips) and 28/07/2012 (Saturday; 40,990 trips)

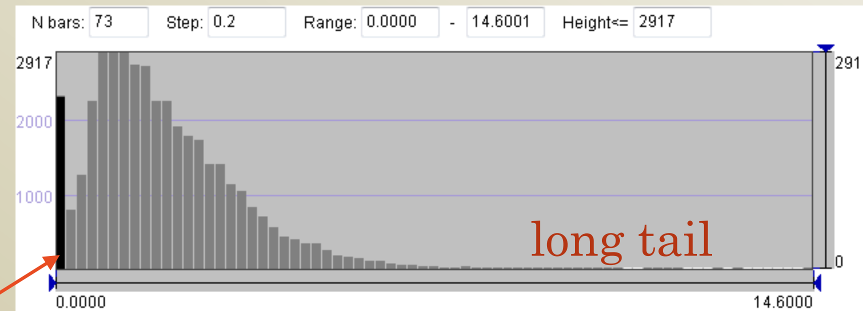


OD moves as generic objects

Examples of analysing the attribute value distributions

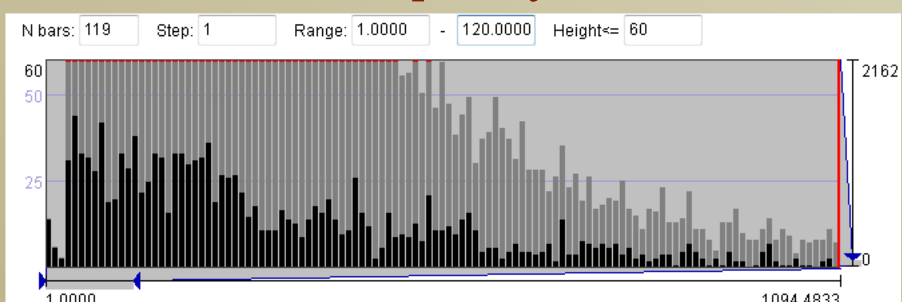
Frequency distributions of the bike **displacement distances** on 25/07/2012 (Wednesday; 41,380 trips) and 28/07/2012 (Saturday; 40,990 trips)

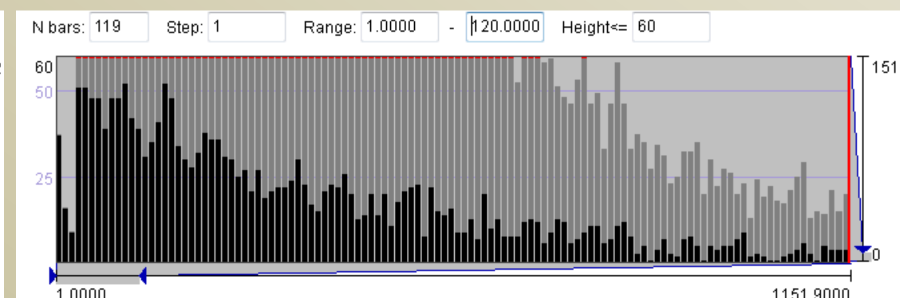




Round trips (possibly, for leisure). It makes sense to analyse them separately.

Frequency distributions of the round trip durations

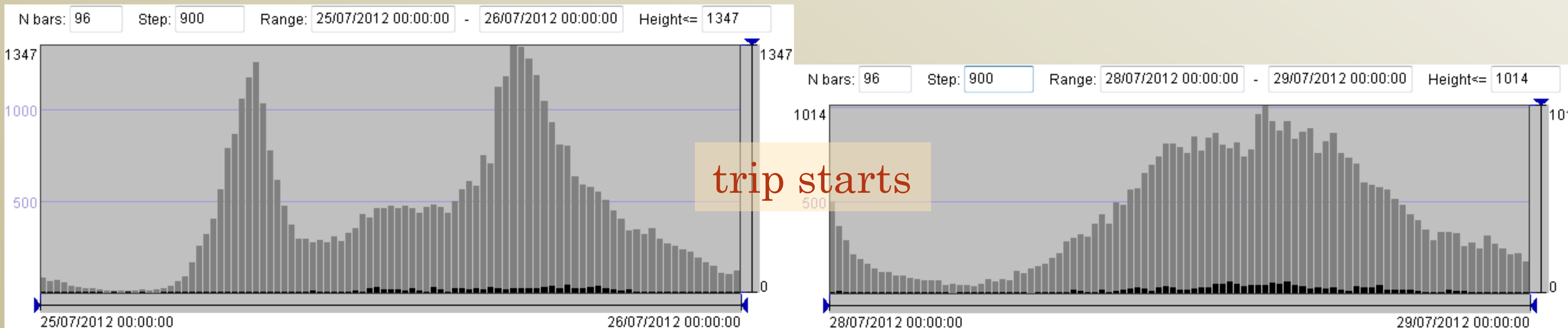




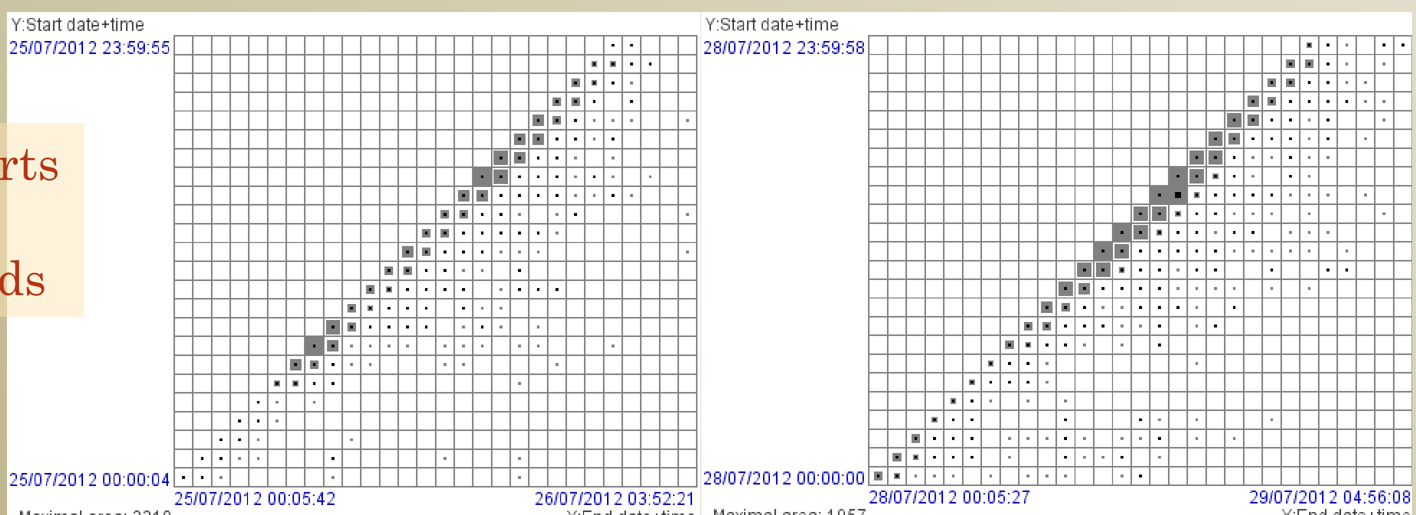
OD moves as temporal objects (events)

Analysis and comparison of temporal distributions

25/07/2012 (Wednesday; 41,380 trips) vs. 28/07/2012 (Saturday; 40,990 trips)

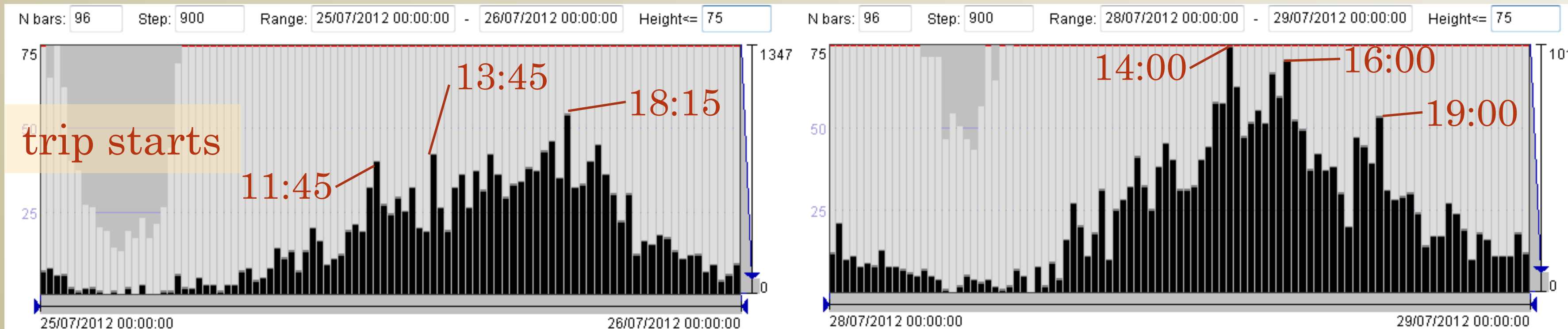


trip starts
vs.
trip ends

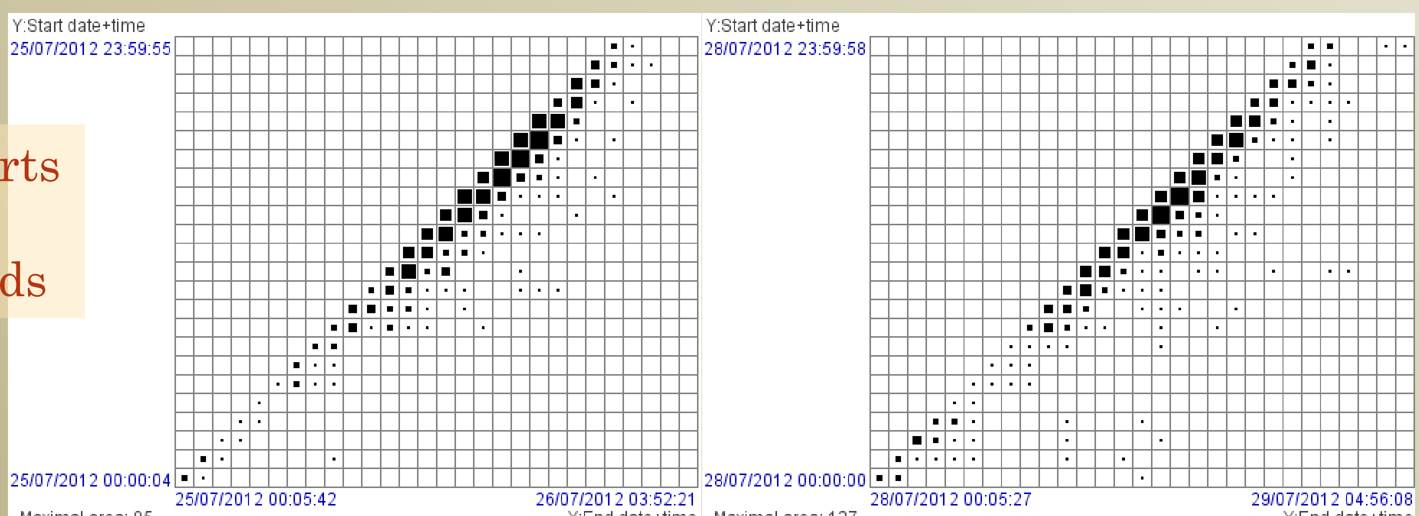


Temporal distributions of the round trips

25/07/2012 (Wednesday; 1,641 trips, 4%) vs. 28/07/2012 (Saturday; 2,322 trips, 5.7%)



trip starts

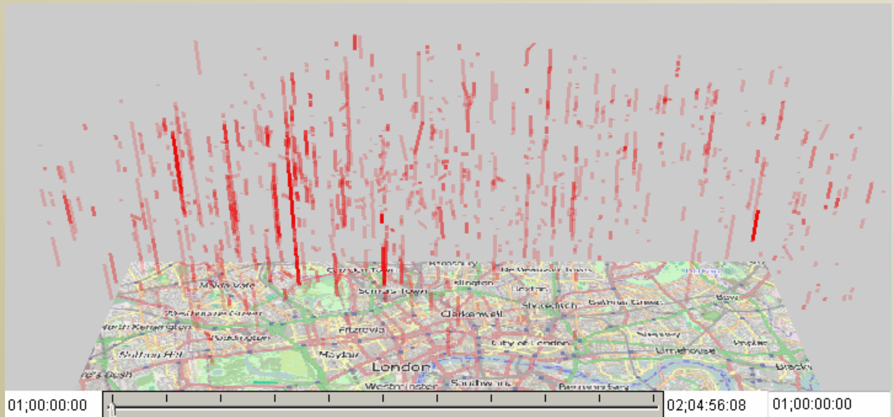


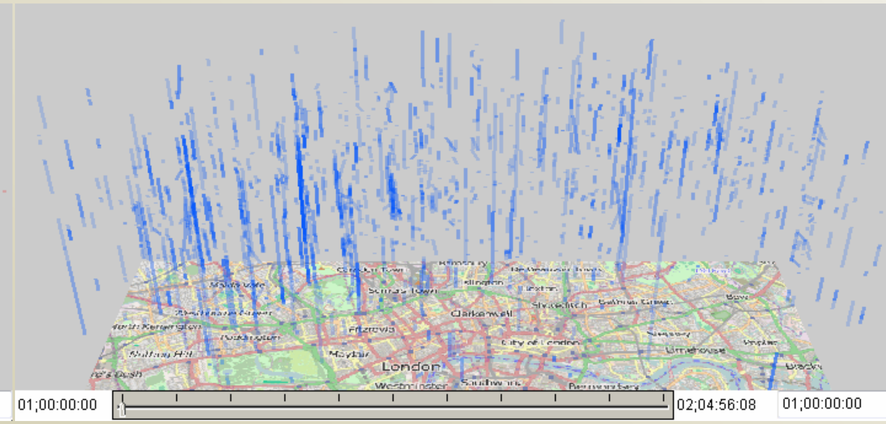
trip starts
vs.
trip ends

Spatio-temporal distributions of the round trips

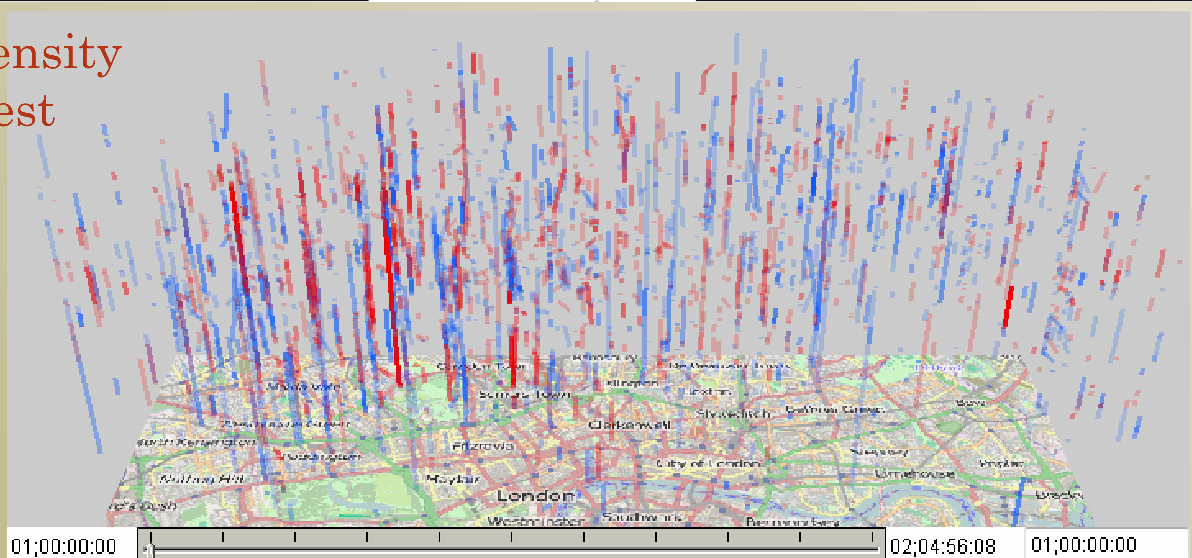
Wednesday

Saturday



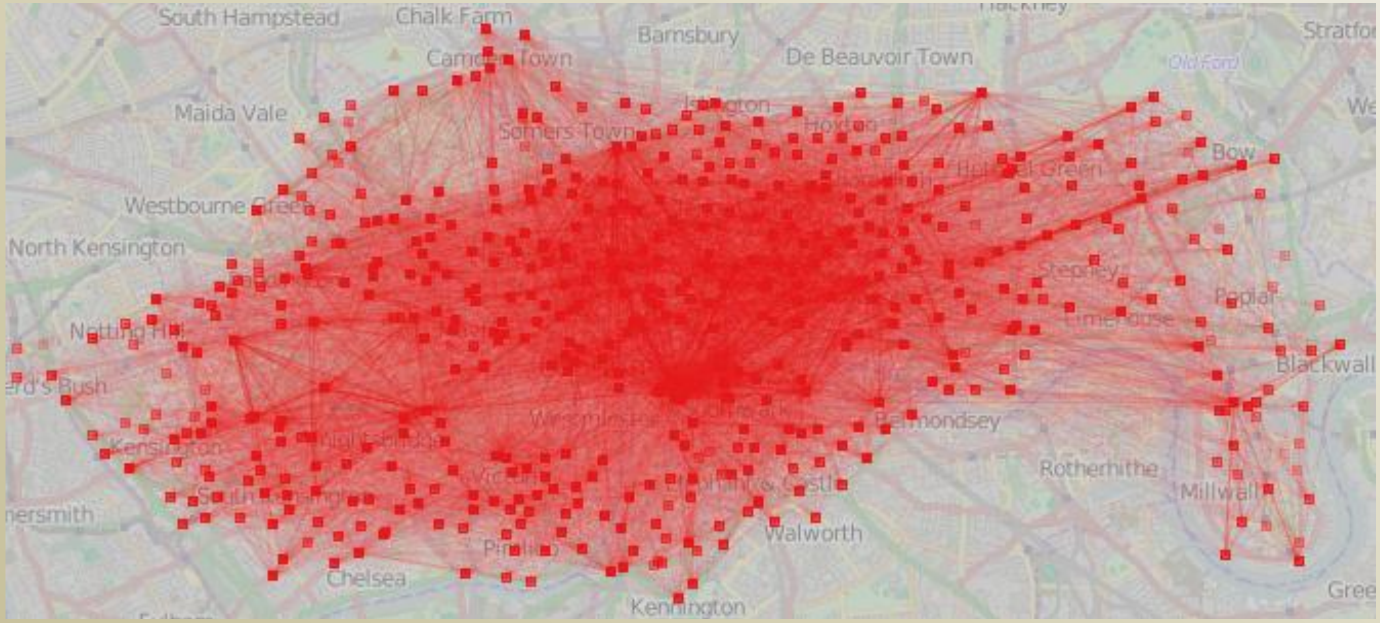


higher density
on the west



Time transformation to the daily cycle

OD moves as spatio-temporal objects are hard to visualise

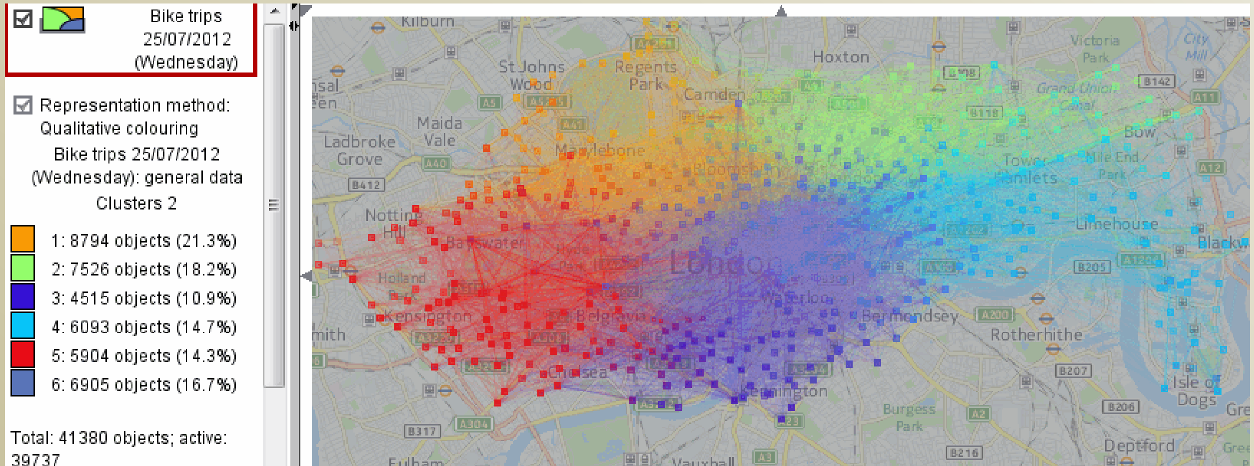


OD moves with differing origins and destinations can be represented by straight lines on a map or in a space-time cube connecting the positions of the starts and ends. However, such a display is very often extremely cluttered due to numerous line intersections.
Here: 39,737 bike trips from Wednesday are shown on a map with 99% transparency.

General approaches to dealing with large data and display clutter

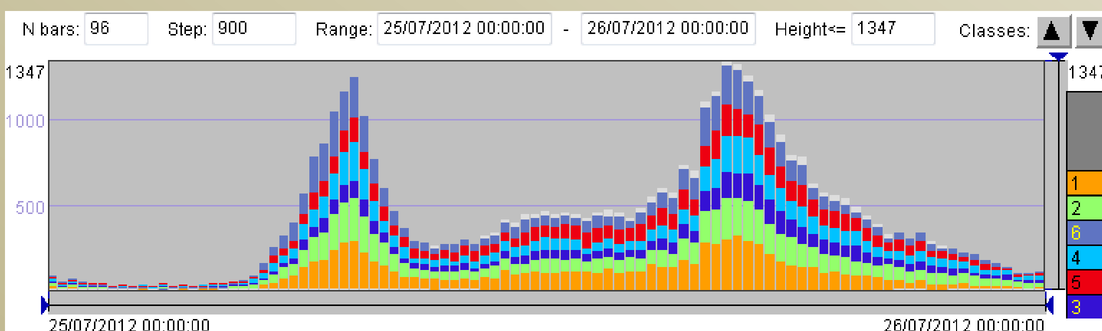
- **Aggregation**: divide attribute domains into bins; obtain counts of objects and summaries of their thematic attributes for the bins; analyse the distributions of the aggregates.
  - In particular, aggregate spatio-temporal objects by areas and time steps.
- **Partition-based clustering**: divide objects into groups by similarity or closeness; analyse and compare group summaries (aggregates) and internal variations.
  - In particular, cluster spatio-temporal objects by their positions in space and time.
  - PBC is efficient when a relatively small number of clusters is enough.
- **Density-based clustering**: find dense groups (concentrations) of similar or close objects; analyse their relations to the remainder.
  - In particular, find spatio-temporal concentrations of spatio-temporal objects; analyse where and when they occurred.

PBC by spatial positions (starts + ends)



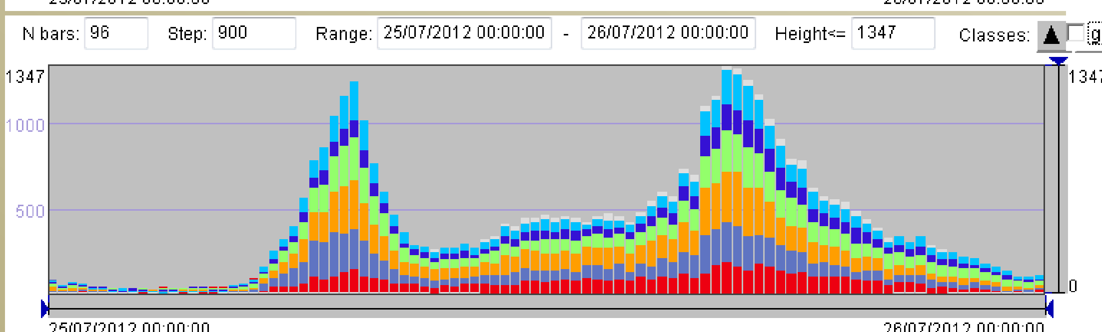
the round trips are excluded

Problem: when clusters are few, the internal variance is very high.



	N members	Track length, mean	Track length, std.dev.	Track length, median	Duration (minutes), mean	Duration (minutes), std.dev.	Duration (minutes), median
1	8794	2.256	1.3507	1.96	21.2	27.55	15
2	7526	2.127	1.3384	1.75	17.8	30.15	13
6	6905	2.384	1.4515	2.05	20.9	28.27	15
4	6093	2.273	1.4104	1.90	17.5	19.49	14
5	5904	2.194	1.2975	1.92	26.0	43.11	17
3	4515	2.316	1.5040	1.95	21.6	40.71	15

group by classes   Sort by: descending   TableLens   condensed   Attribute...



DBC: what distance function to use?

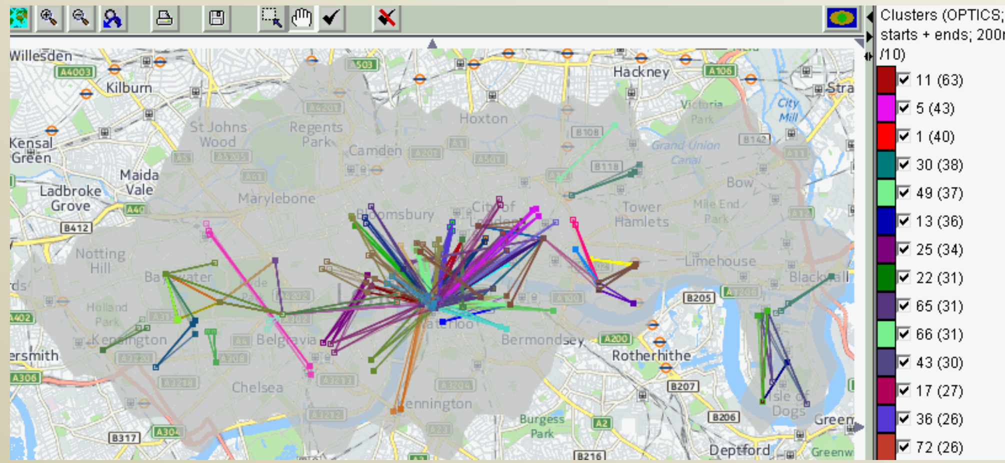
- DBC requires setting a distance threshold ⇒ there should be a meaningful notion of the distance between objects.
- The **spatial distance** between OD moves can be defined as the mean of the spatial distances between the origins and between the destinations:
  - $s\_distance(m_1,m_2) = (s\_distance(o_1,o_2) + s\_distance(d_1,d_2)) / 2$
- Analogously, the **temporal distance** may be defined as
  - $t\_distance(m_1,m_2) = (t\_distance(o_1,o_2) + t\_distance(d_1,d_2)) / 2$
- The **spatio-temporal distance** between OD moves can be defined taking the same approach as for instant spatial events:
  - Spatial distance threshold (radius) R_S + temporal distance threshold R_T
  - $s\_t\_distance = \max(s\_distance/R_S + t\_distance/R_T)*R_S$

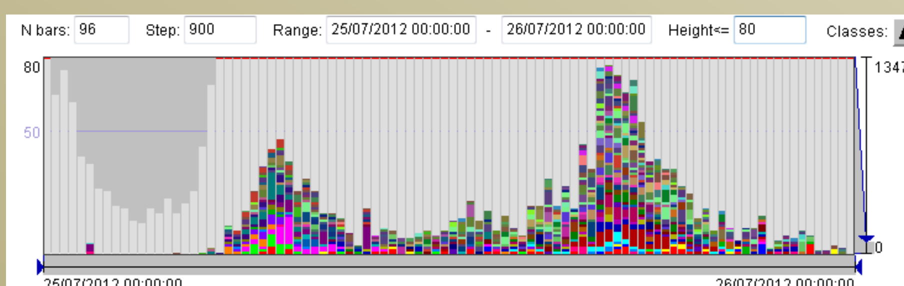
Distance functions and analysis tasks

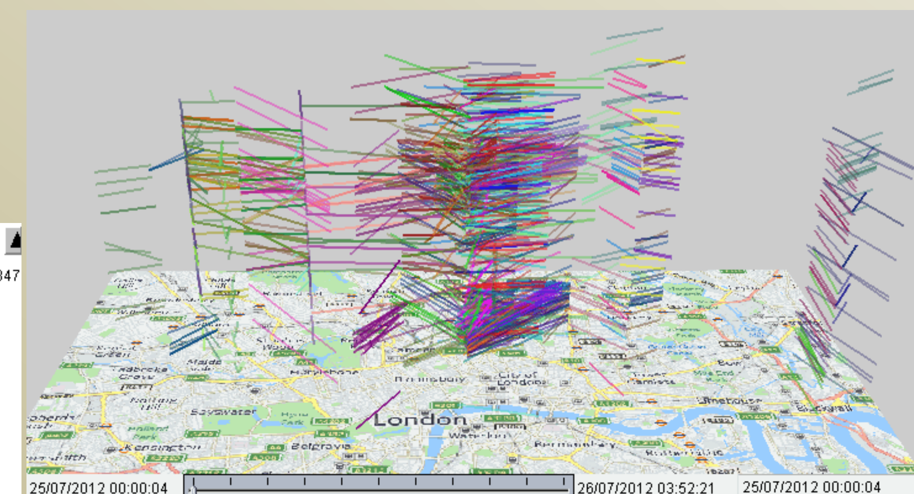
- DBC with different distance functions is used for answering different questions.
- **Spatial distance**: What moves (in terms of the origin and destination *regions*) were the most frequent throughout the whole time?
  - Further questions: What is the proportion of the frequent moves in the whole set? How are the frequent moves distributed over time? What are their thematic characteristics?
- **Spatio-temporal distance**: What similar trips occurred in close times?
  - A spatio-temporal cluster means collective movement of multiple objects between some origin and destination regions.
  - Further questions: How frequent are the occurrences of collective movement? Based on their positions in space and time, how can they be interpreted?

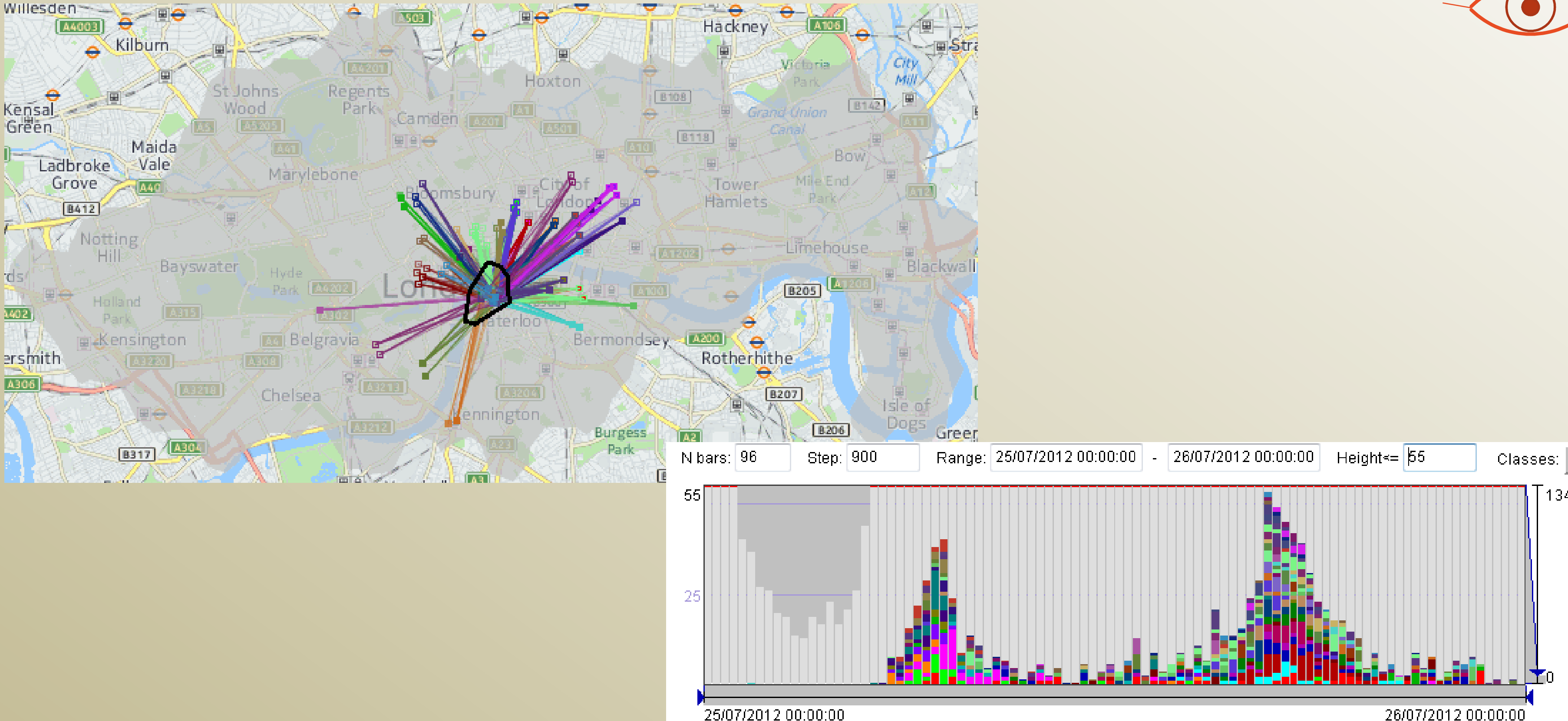
DBC by the spatial distance (examples)

25/07/2012 (Wednesday); round trips excluded.
Distance threshold R = 200 m; minimal number of neighbours N = 10.
Result: 93 spatial clusters with sizes from 10 to 63 include 1,618 trips (4%); the noise consists of 38,119 trips (96%).
Hence, not many very similar trips occurred throughout the day.



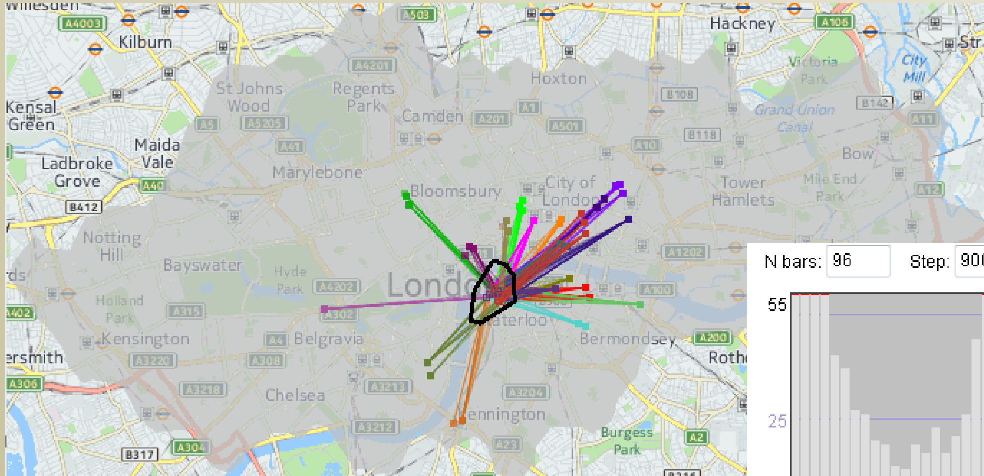




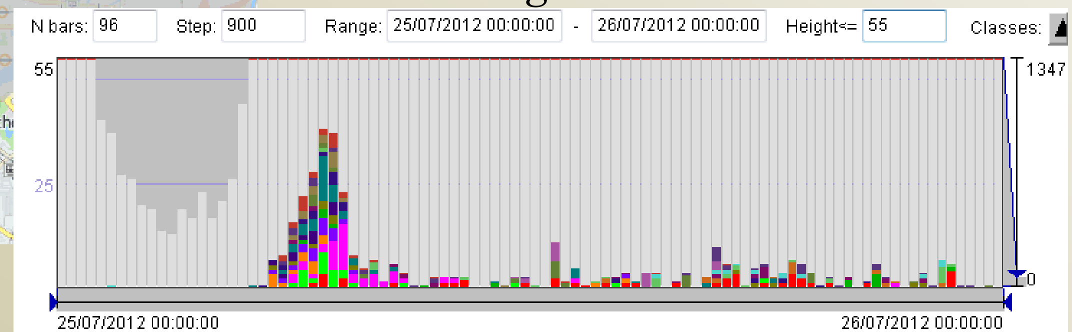


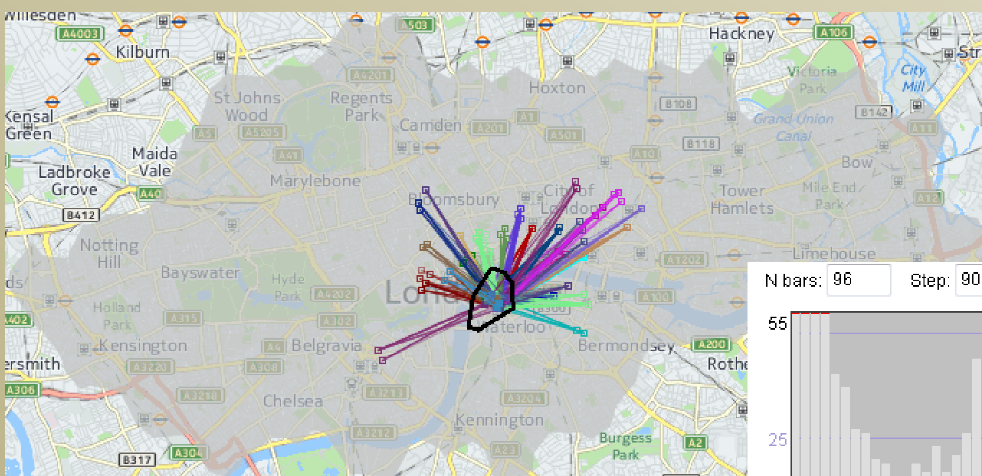
Further observations:

- Waterloo is the most popular area of trip origins and destinations: 940 out of 1,618 clustered trips (58%) originate from or end in this area.
- Some spatial clusters are related to particular time periods, i.e., either morning or evening.

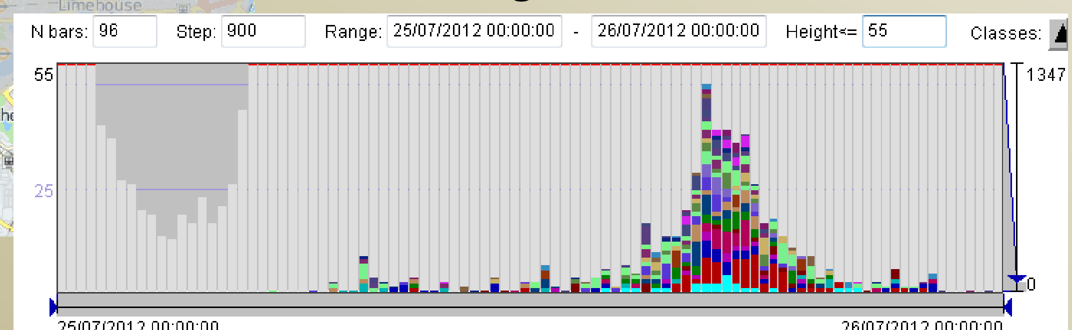


Out-trips from Waterloo:
19 clusters containing 390 trips.
Most of the out trips occurred in the morning.





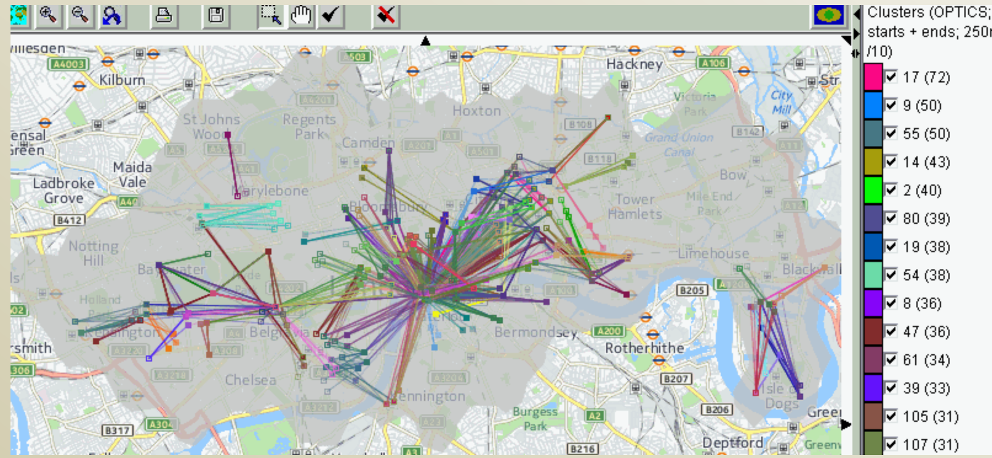
In-trips to Waterloo:
26 clusters containing 550 trips.
Most of the out trips occurred in the evening.

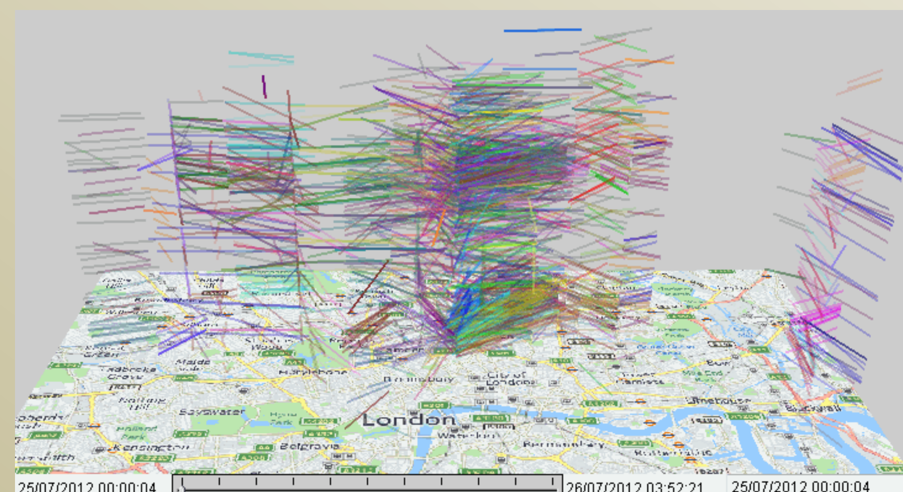


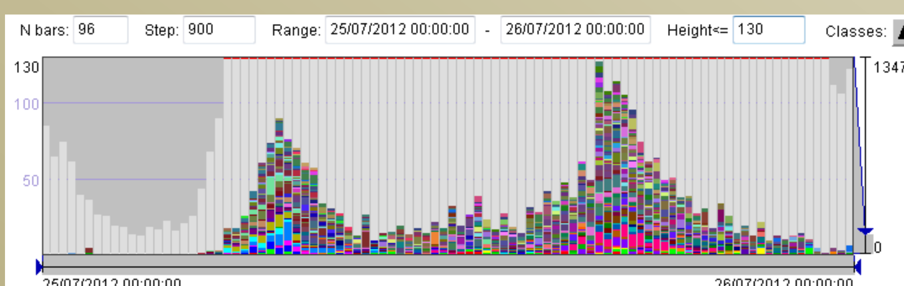
DBC by spatial distance

Testing other parameter settings

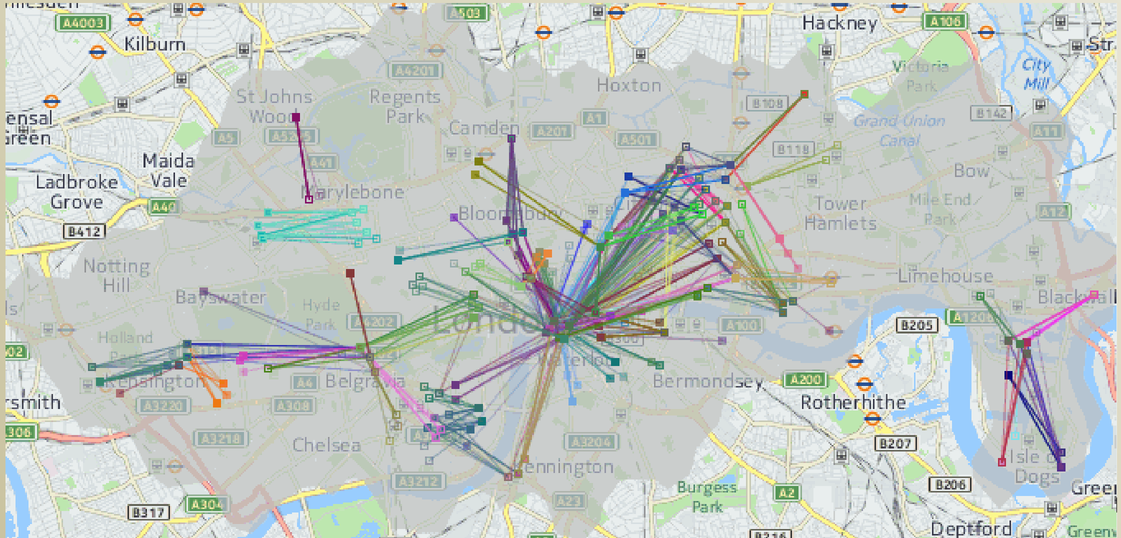
Distance threshold R = 250 m; minimal number of neighbours N = 10.
Result: 153 spatial clusters with sizes from 10 to 72 include 2,707 trips (6.8%); the noise consists of 37,030 trips (93.2%).
Relaxing the notion of similarity increased the proportion of clustered trips.



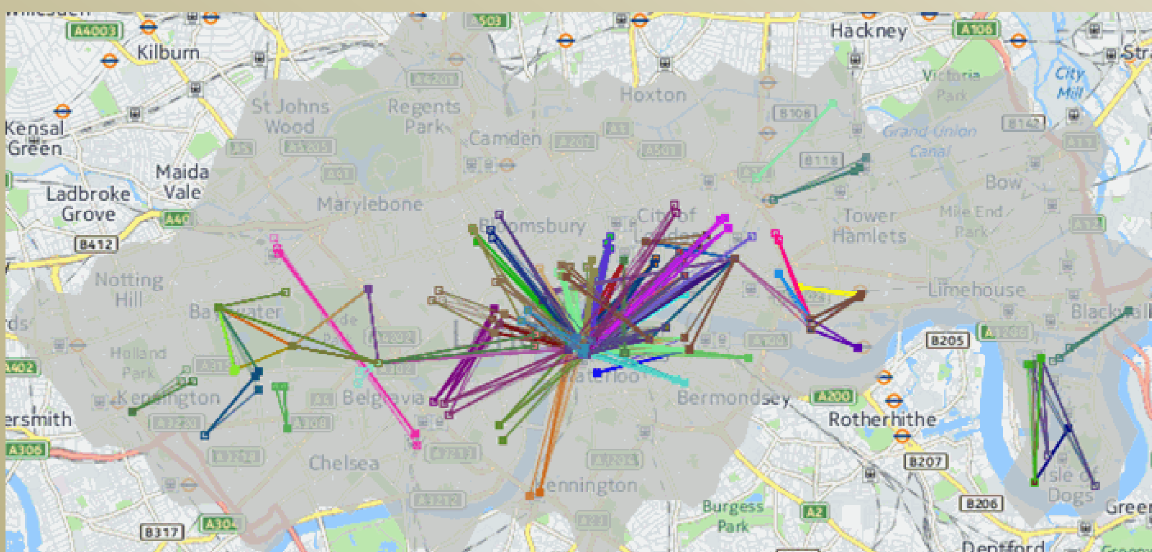




Clustered trips that were previously in the “noise”:

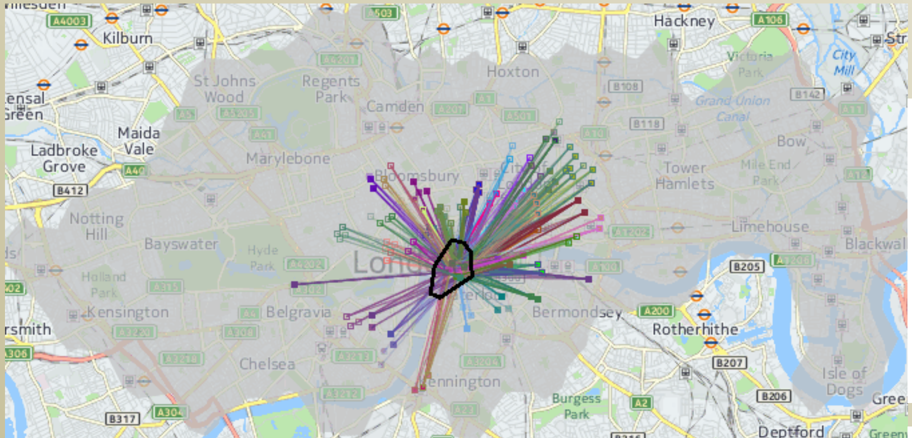


Compare to the previously obtained clusters with R = 200 m:



Evidently, some of the trips from the former “noise” have been grouped together with previously clustered trips; i.e., the previously existing clusters have become bigger. Besides, a number of new clusters have appeared.

Consistency between the clustering results



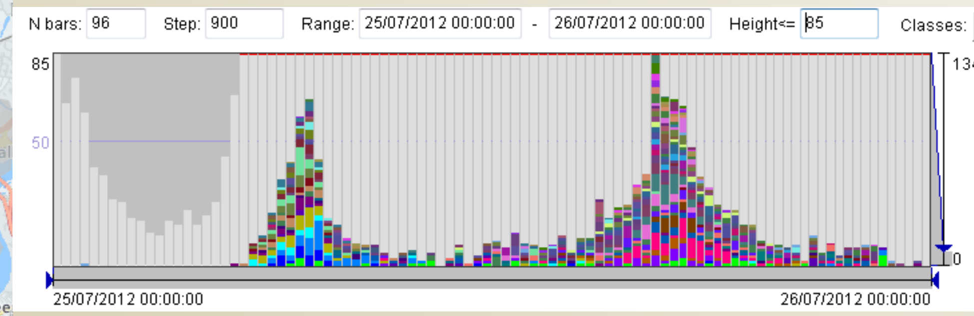
Popularity of the Waterloo area:

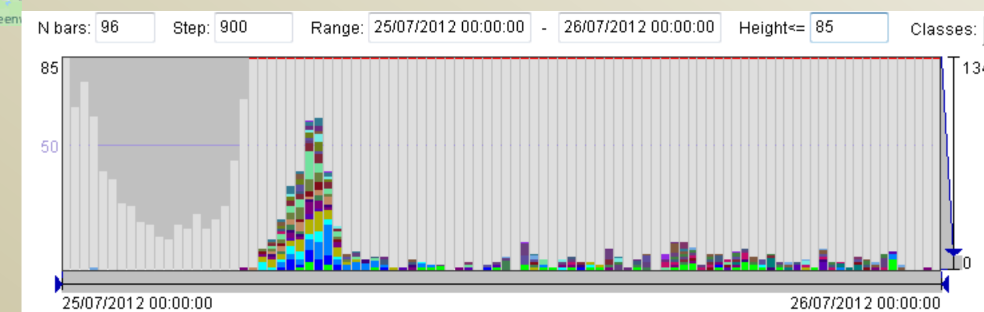
58 spatial clusters with 1,358 outgoing and incoming trips. The trip number increased by 44% from previous 940, but the proportion among the clustered trips decreased to 50% from the previous 58%.

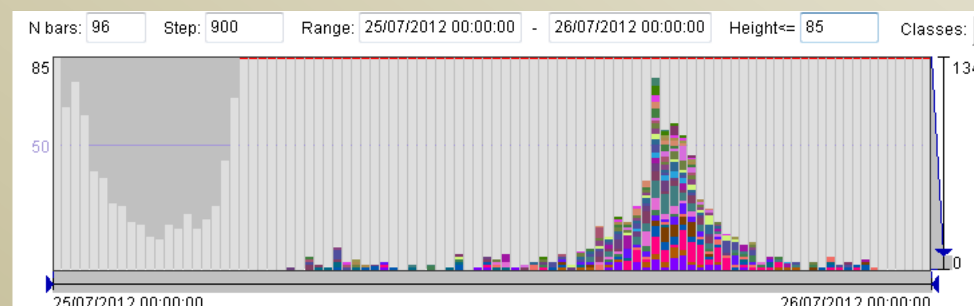
Outgoing trips: 29 clusters with 627 trips; the majority occurred in the morning.

Incoming trips: 30 clusters with 742 trips; the majority occurred in the evening.

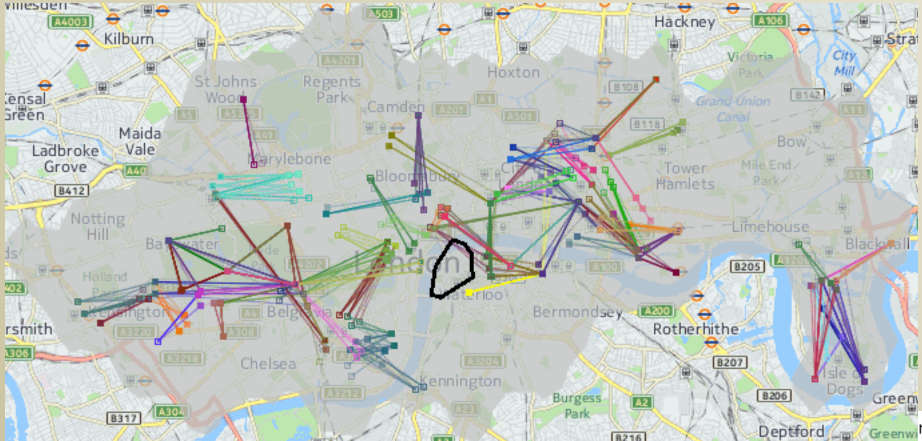
There were 11 trips that both started and ended in the Waterloo area.



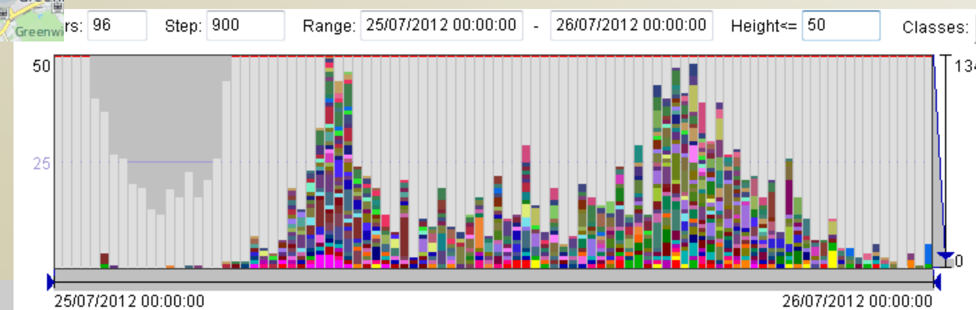


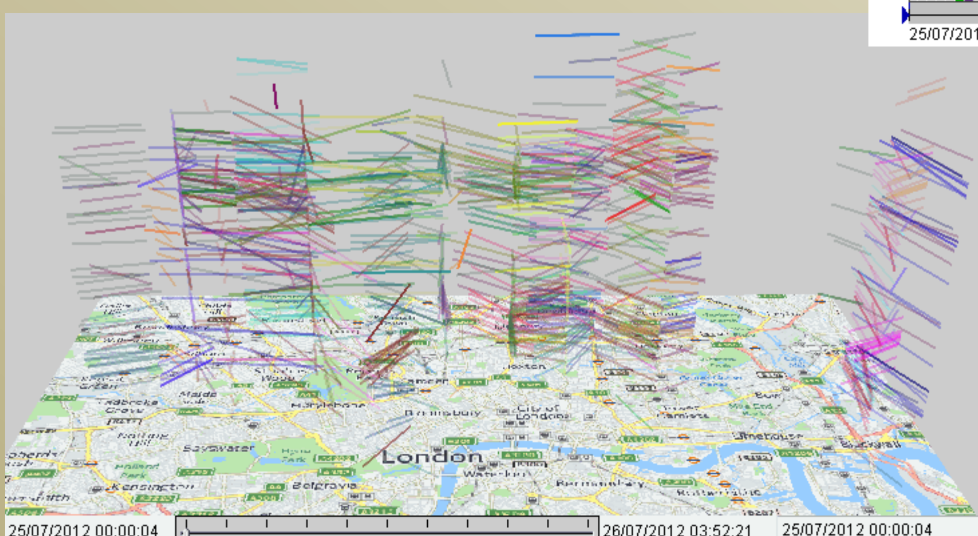


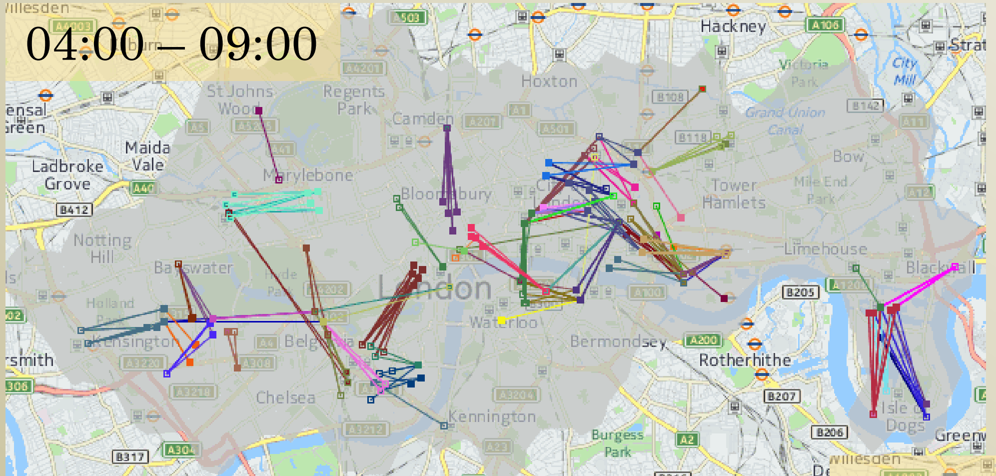
The trip clusters outside of the Waterloo area:



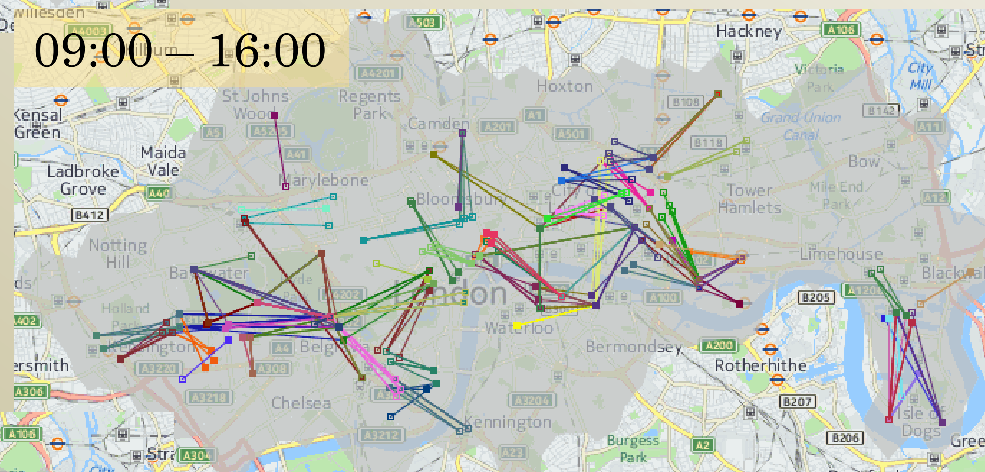
94 clusters; 1,349 trips

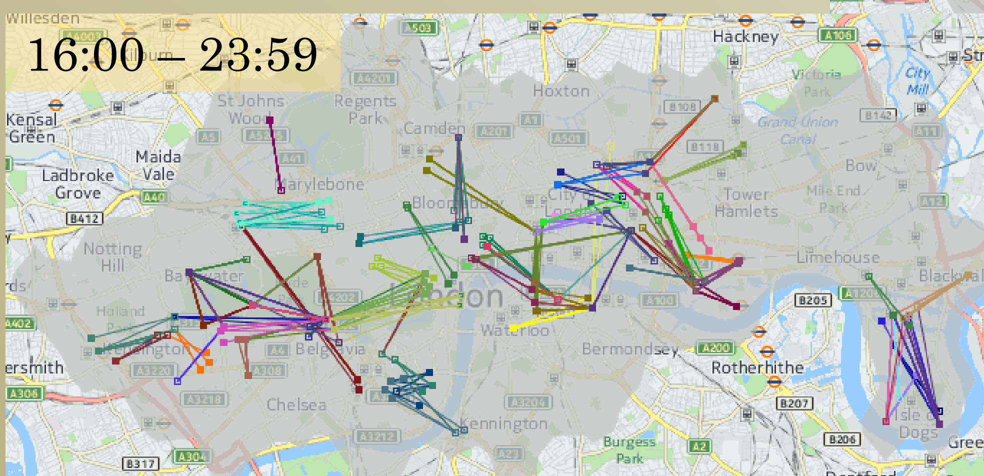






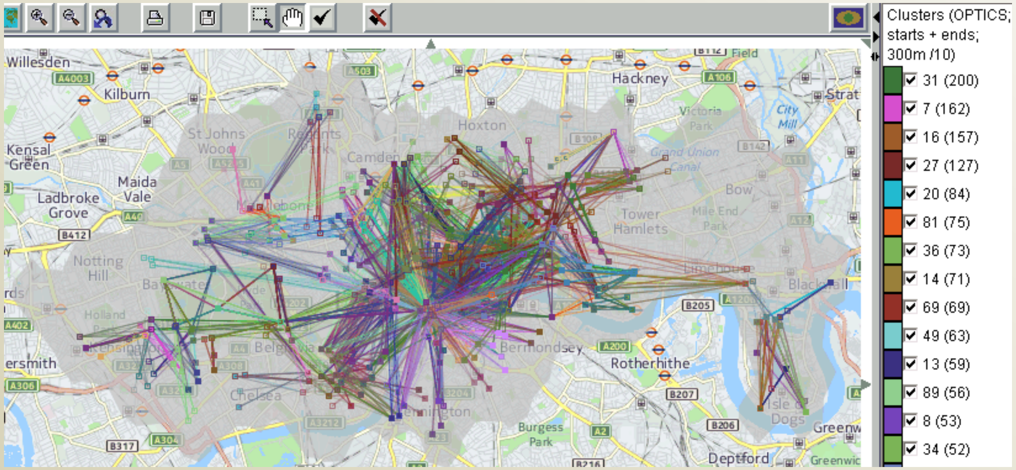
We can notice cluster pairs with opposite movements. In some of these pairs, one cluster mostly appears in the morning and the other in the evening.

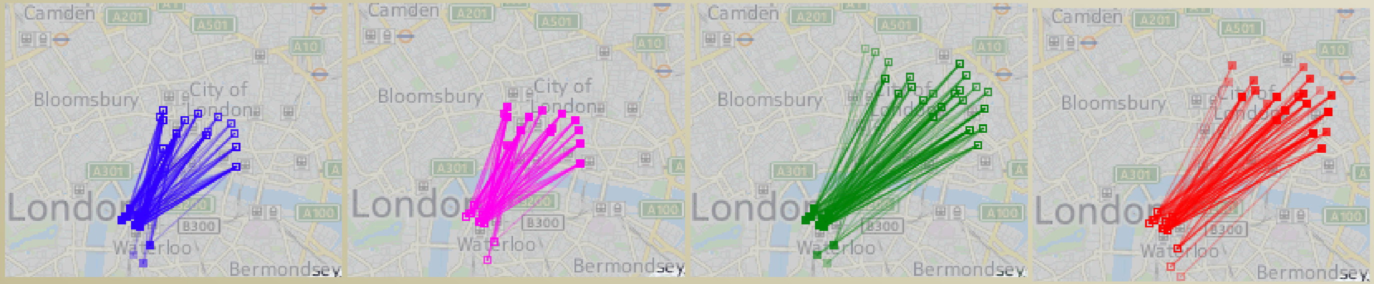




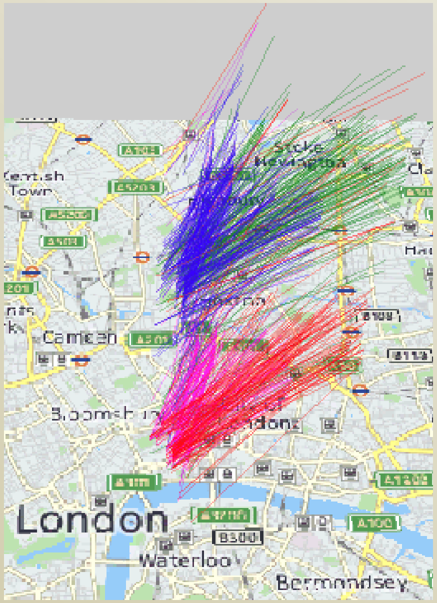
However, there is no such vivid separation between the morning and evening movement directions as for the Waterloo area.

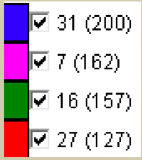
Distance threshold R = 300 m; minimal number of neighbours N = 10.
Result: 277 spatial clusters with sizes from 10 to 200 include 6,174 trips (15.5%); the noise consists of 33,563 trips (84.5%).





The further relaxation of the notion of similarity increased much more the proportion of clustered trips. However, we have obtained many fuzzy, untidy clusters. For example, the four largest clusters have too high spatial dispersion of the trip origins or destinations.

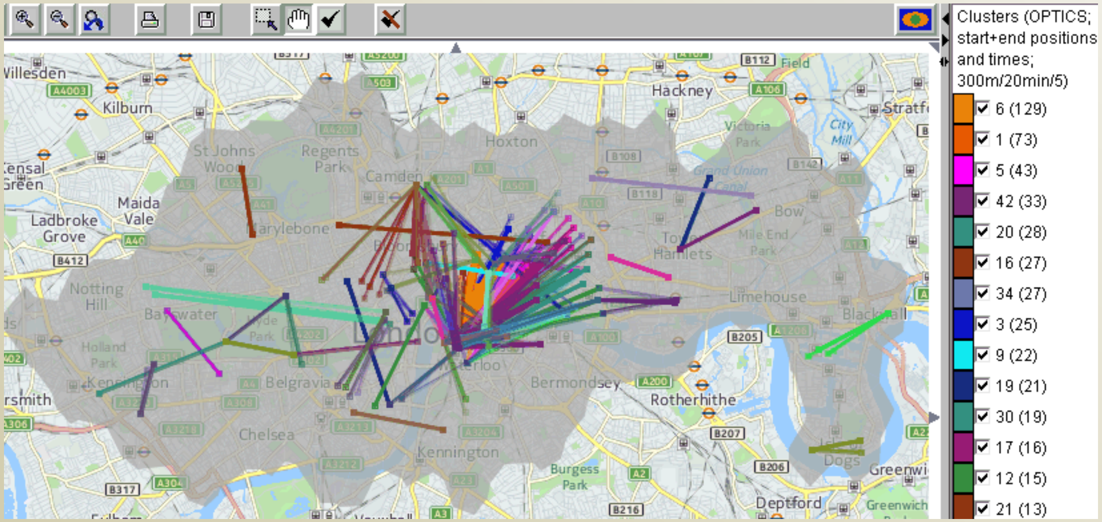


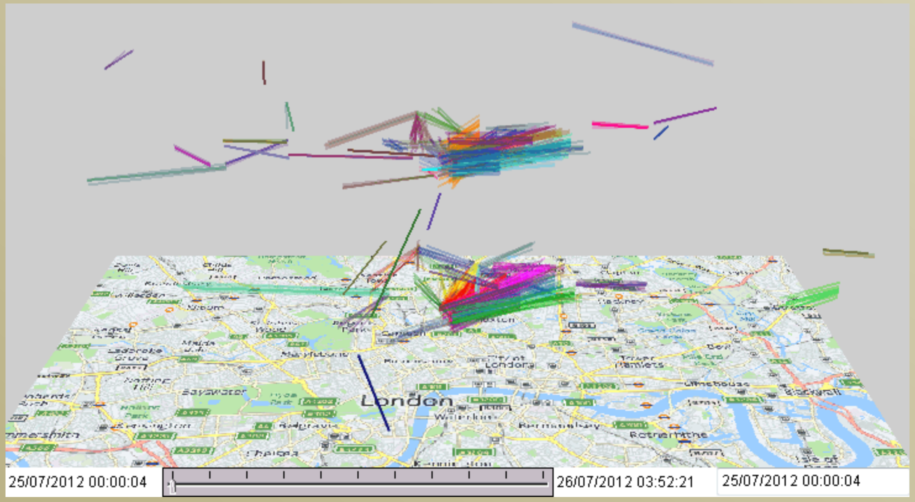


What similar trips occurred in close times?

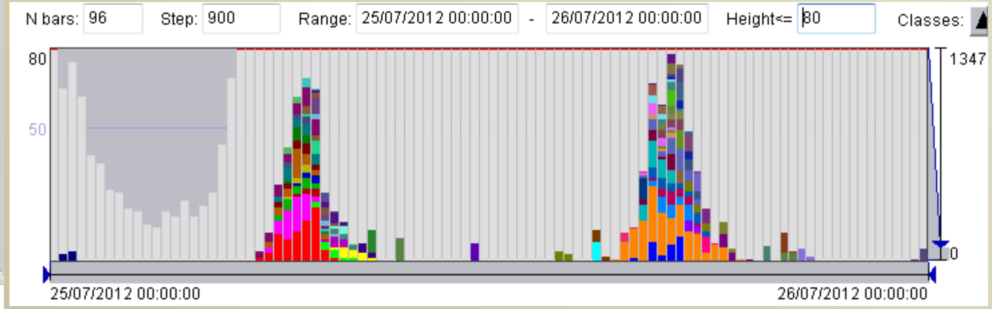
Distance thresholds R_{space} = 300 m and R_{time} = 20 minutes; minimal number of neighbours N = 5.

Result: 75 spatio-temporal clusters with sizes from 5 to 129 include 916 trips (2.3%).

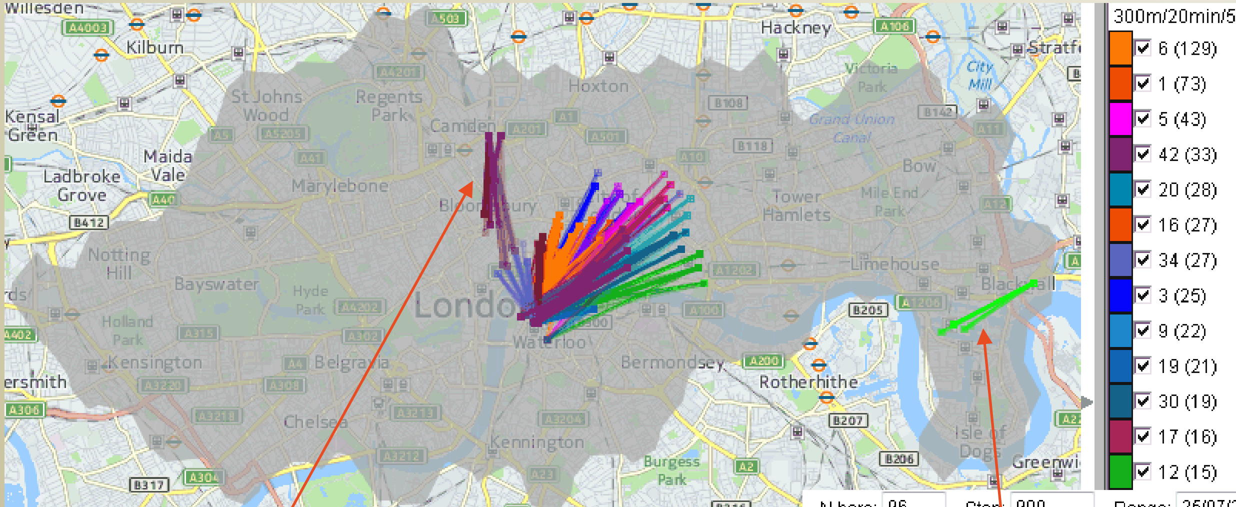




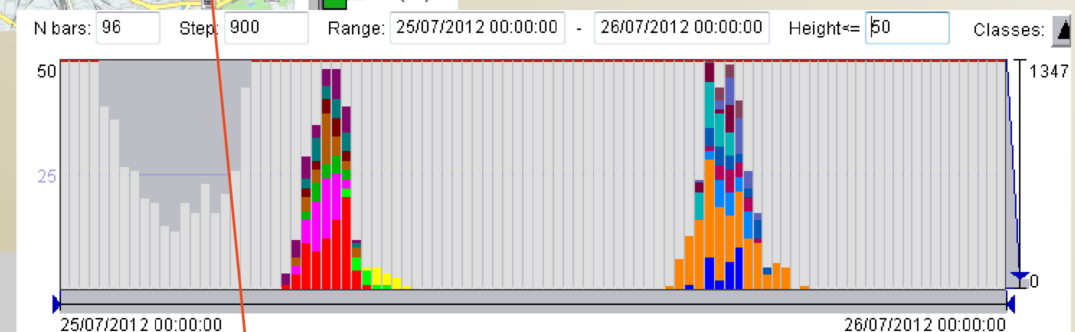
Groups of similar trips (collective or mass movements) mostly occurred in the morning and in the evening.



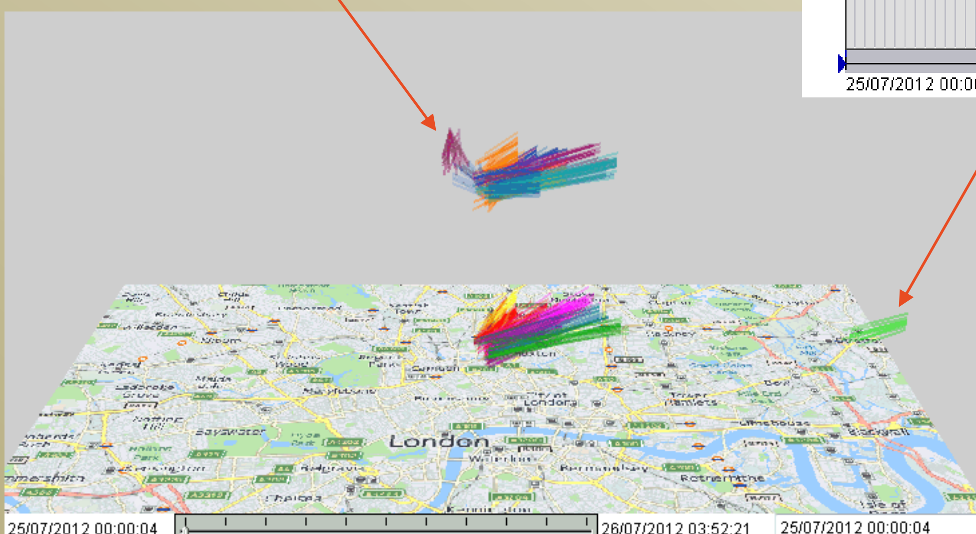
Groups with ≥ 10 members



18 groups with 536 members in total.
All occurred in the morning or in the evening.
All but 3 groups have origins or destinations in the Waterloo area.

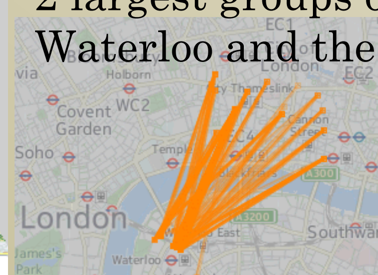


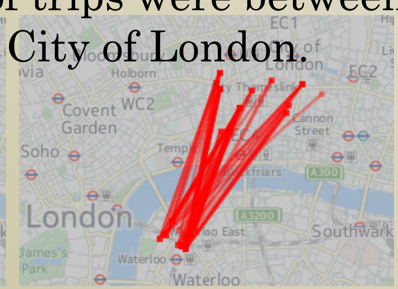
Two groups with 13 and 10 trips end near King's Cross in the evening.



One group consists of 10 morning trips from East India to Canary Wharf.

2 largest groups of trips were between Waterloo and the City of London.





Analysis of OD moves using spatial clustering

Iterative clustering + visualisation+ interactive operations

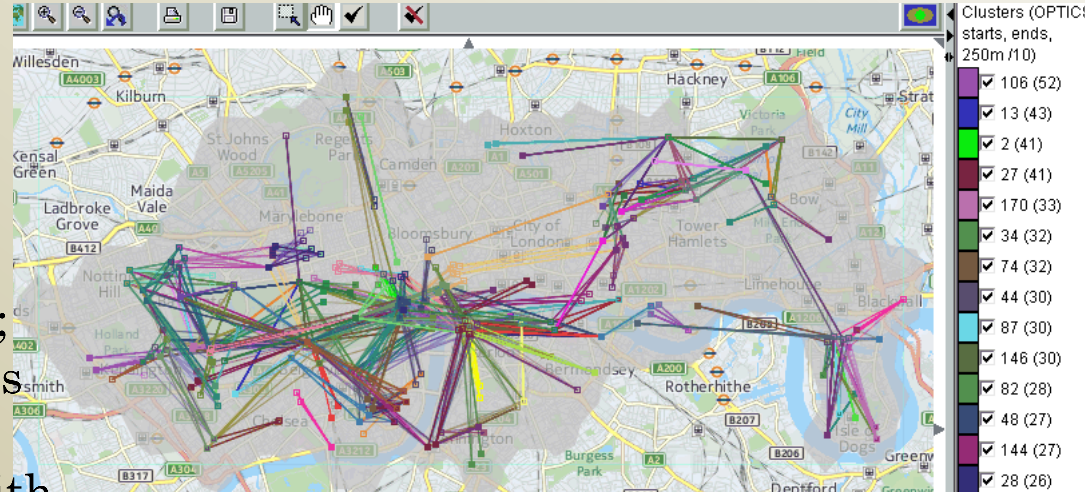
- In this example, we applied the principles of visual analytics:
  - Iterative application of computational methods (here clustering) with different parameter settings.
  - Visual investigation of the results, comparison between them.
  - Division of data into subsets and comparison between the subsets.
- In the investigation, we used following interactive operations:
  - Display linking by highlighting
  - Filtering by attribute values (e.g., displacement distance)
  - Propagation of cluster colours
  - Filtering by cluster selection
  - Spatial filtering by visited areas
  - Temporal filtering
- As a result, we gained useful information from quite a large dataset despite of initially extremely cluttered views.

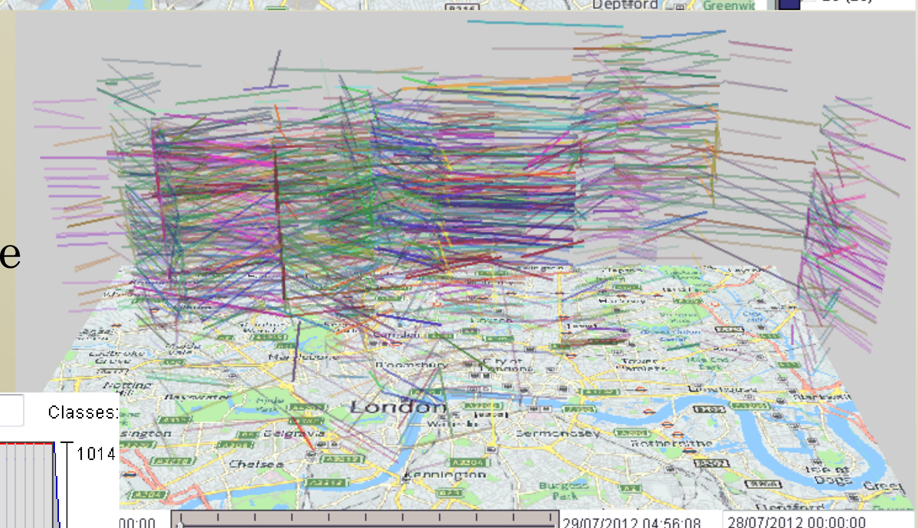
What we have learned so far

- High variety of trips; not many similar trips.
- A large number of similar trips from Waterloo in the morning and to Waterloo in the evening.
  - These may be trips of commuters who travel to London by railway and then get to the final destinations by bike.
- Existence of opposite groups of trips.
  - Some of these may be trips of people going to their work or study places in the morning and back in the evening.
- Existence of groups of similar trips occurring close in time.
  - Most such groups occur in the morning or in the evening and go from or to areas around railway stations.

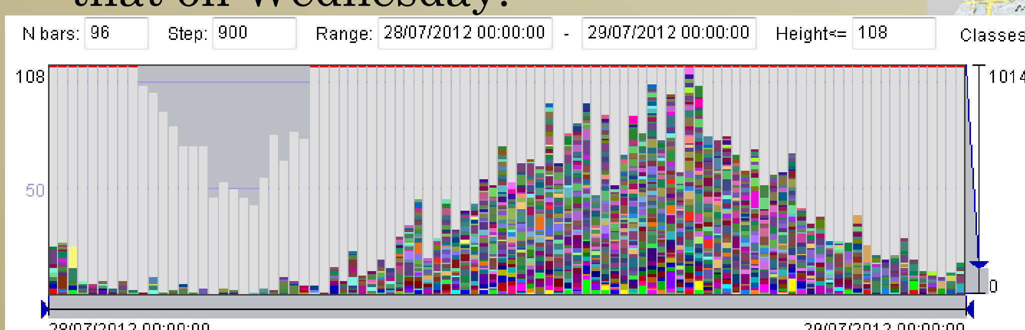
Comparison of ST behaviours in different days *(very briefly)*

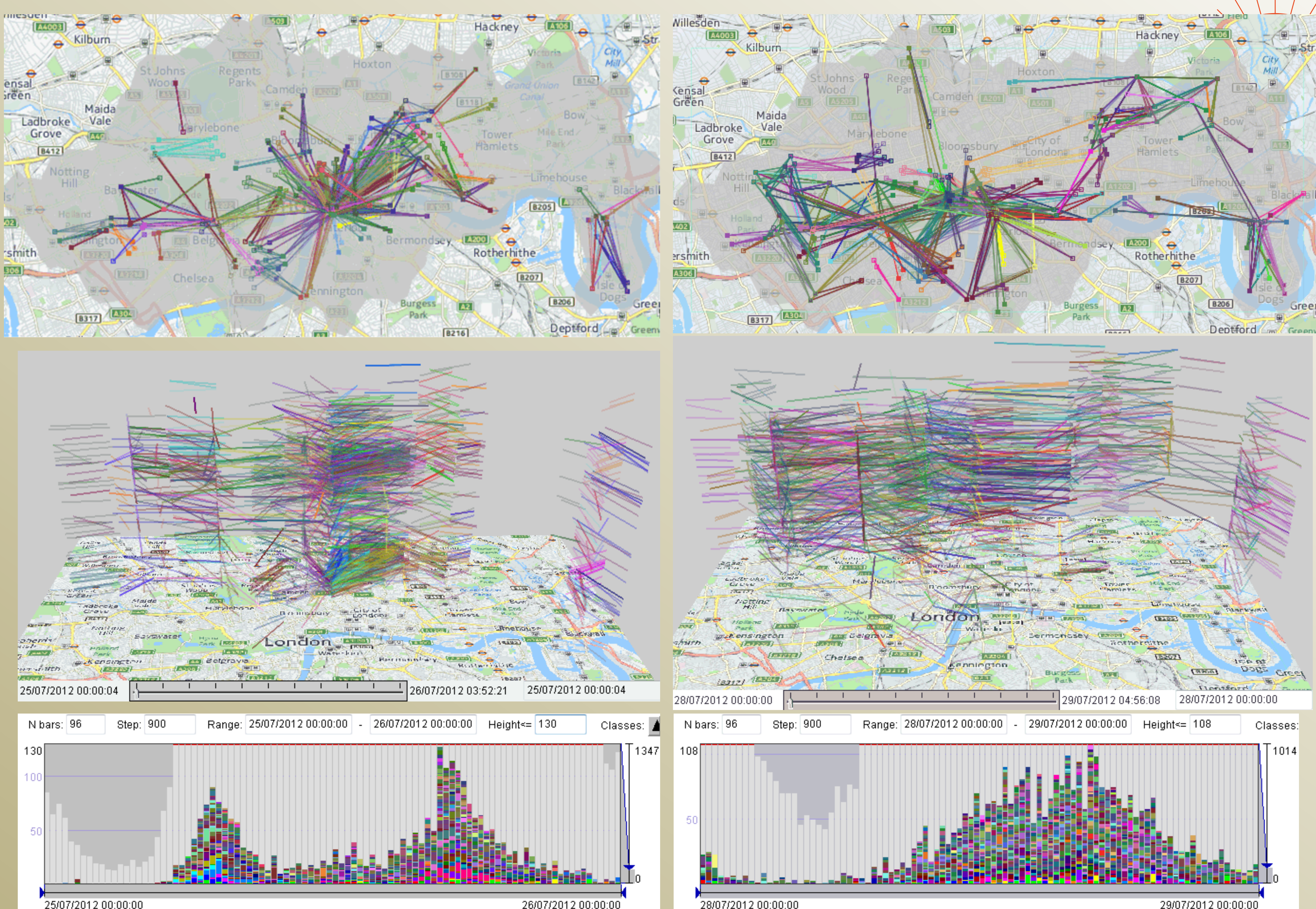
28/07/2012 (Saturday); round trips excluded, 38,668 trips remain.
Distance threshold R = 250 m; minimal number of neighbours N = 10.
Result: 212 spatial clusters with sizes from 10 to 52 include 3,399 trips (9.2%); the noise consists of 35,119 trips (90.8%).





The spatio-temporal distribution of the Saturday trips differs greatly from that on Wednesday.




25/07/2012 00:00:04
26/07/2012 03:52:21
25/07/2012 00:00:04
N bars: 96
Step: 900
Range: 25/07/2012 00:00:00 - 26/07/2012 00:00:00
Height<= 130
Classes:
130
100
50
1347
0
25/07/2012 00:00:00
26/07/2012 00:00:00
28/07/2012 00:00:00
29/07/2012 04:56:08
28/07/2012 00:00:00
N bars: 96
Step: 900
Range: 28/07/2012 00:00:00 - 29/07/2012 00:00:00
Height<= 108
Classes:
108
50
1014
0
28/07/2012 00:00:00
29/07/2012 00:00:00

Questions?

Analysis of OD movement data by means of clustering

Analysis of OD movement data by means of aggregation

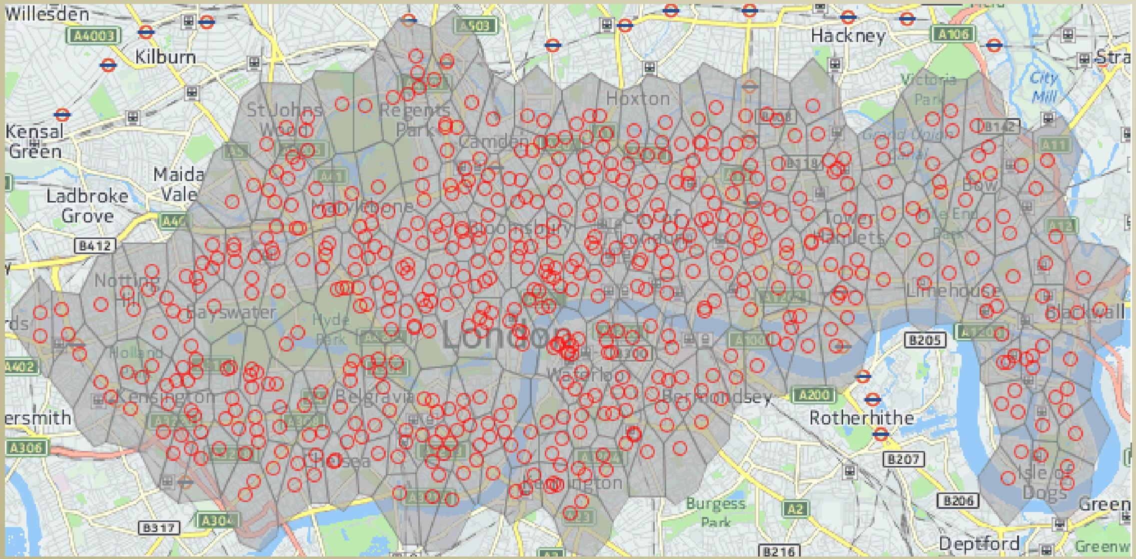
Involves data transformation to spatial time series

Spatio-temporal aggregation of spatio-temporal objects (events)

- The space (territory) containing the events is divided into compartments.
- The time span of the data is divided into intervals, usually of equal length.
- For each pair (compartment × time interval), the relevant events are summarized into event counts + summaries of thematic attributes
  - Relevant **OD moves** for a pair (compartment × time interval): moves that start in this compartment during this time interval + moves that end in this compartment during this time interval
    - The starts and ends are counted separately.
- Aggregation result: spatial time series A(S,T)
  - S: the set of space compartments
  - T: the set of time intervals
  - A: aggregate attributes (counts + statistical summaries of thematic attributes)

Space (territory) division possibilities

- Regular grid (rectangular or hexagonal) with a chosen cell size
- <u>Meaningful</u> pre-existing division (e.g., administrative units)
- Irregular grid respecting the spatial distribution of the objects

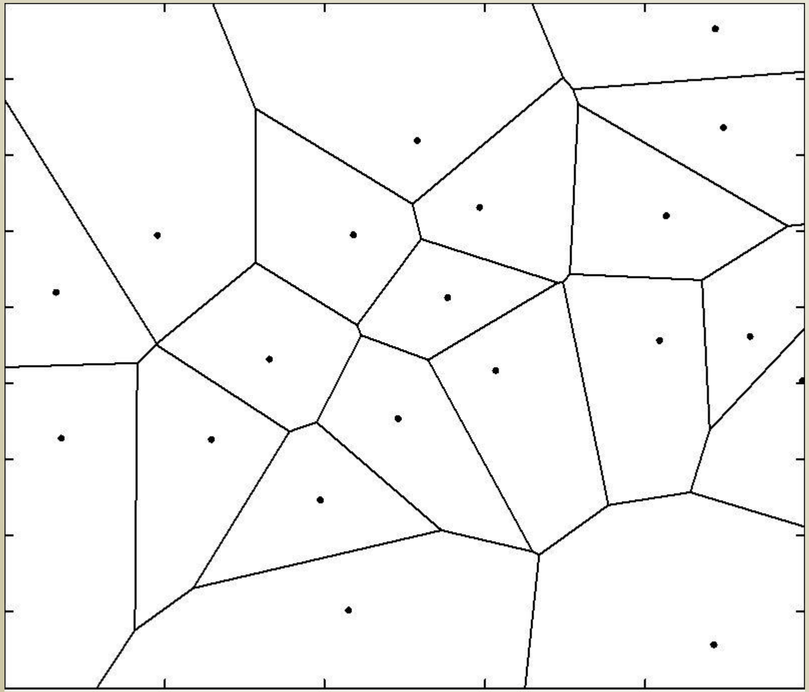


Example: an irregular grid for the London cycle hire data has been built on the basis of the spatial distribution of the docking stations, i.e., the origin and destination places of the trips.

Voronoi tessellation (a.k.a. Voronoi diagram)

Used for building irregular grids

- The partitioning of a plane with N points into convex polygons *(cells),* such that
  - each polygon contains exactly one generating point
  - every point in a given polygon is closer to its generating point than to any other.



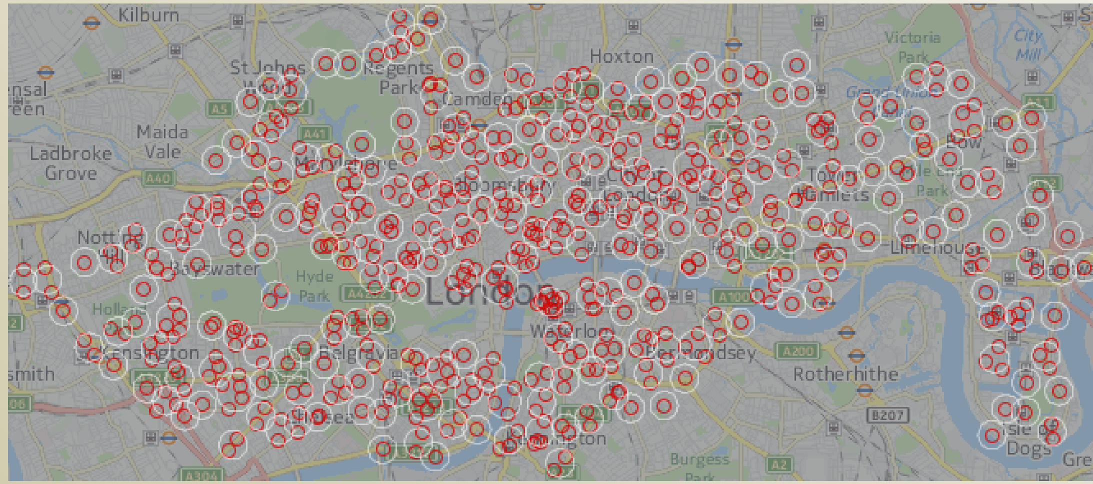
- The generating points are also called *seeds*.
- A Voronoi diagram is also known as a Dirichlet tessellation.
- The cells are called Dirichlet regions, Thiessen polytopes, or Voronoi polygons.

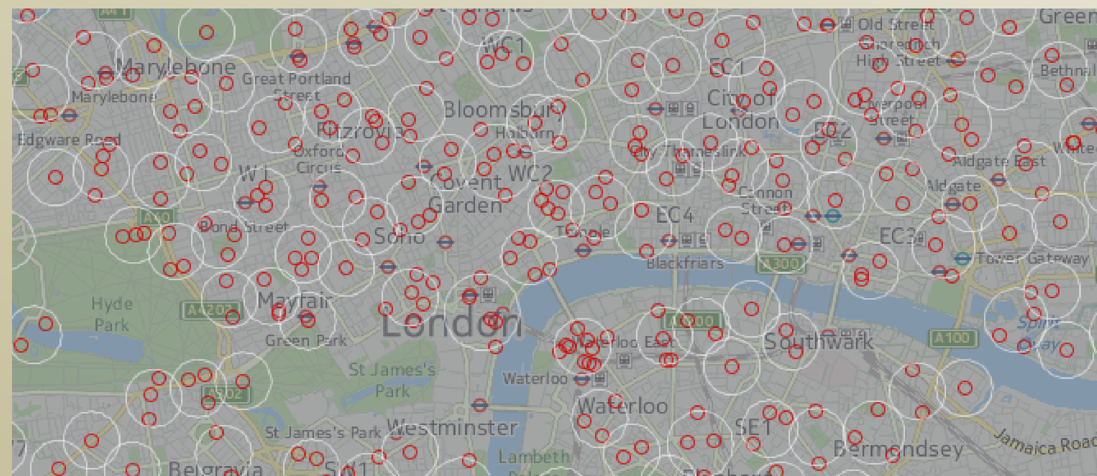
How do we obtain a tessellation?

Step 1: we apply a special algorithm for clustering of points based on their spatial proximity.

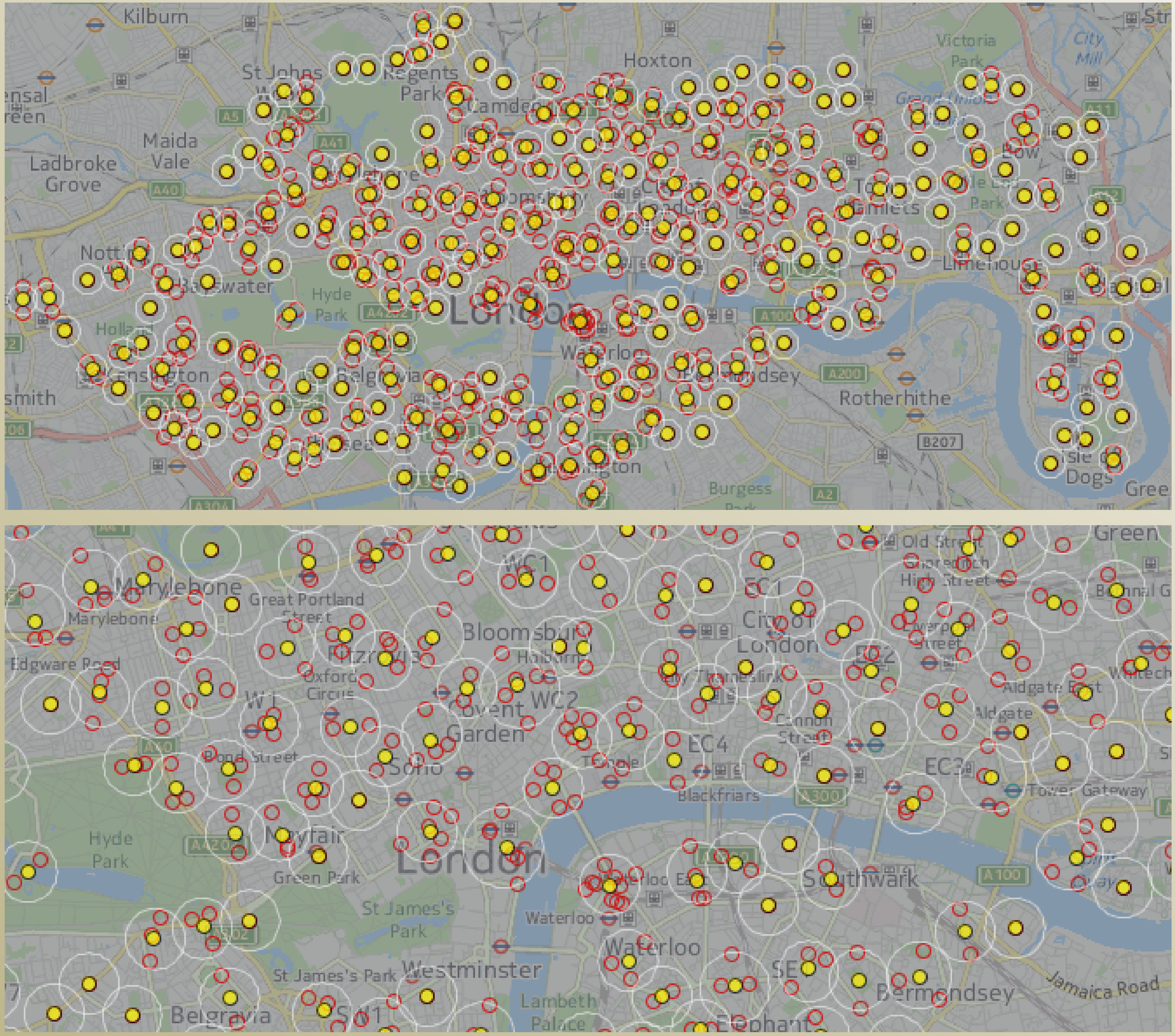
The main idea of the algorithm: put the points into circles with a given maximal radius R.

We chose R = 400 m.



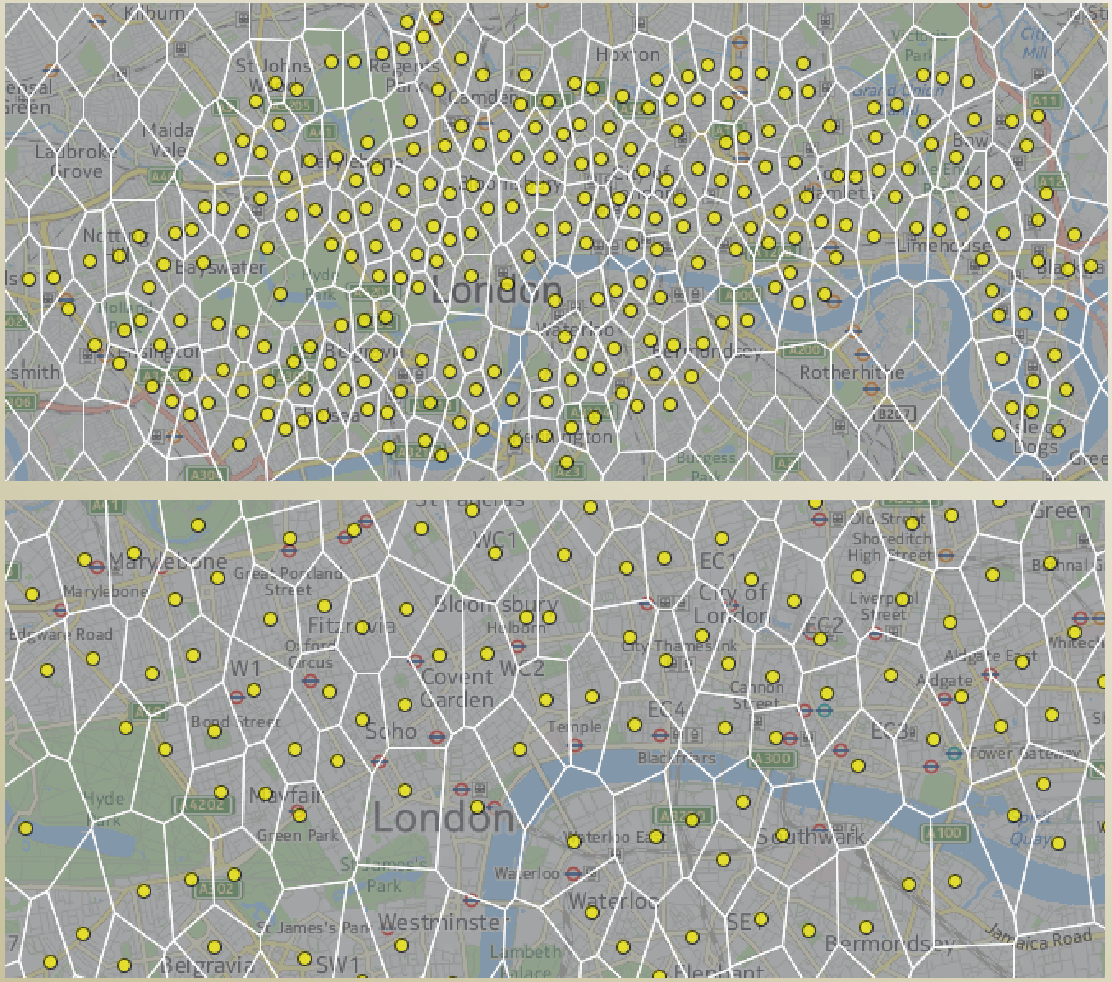


The centres of the circles

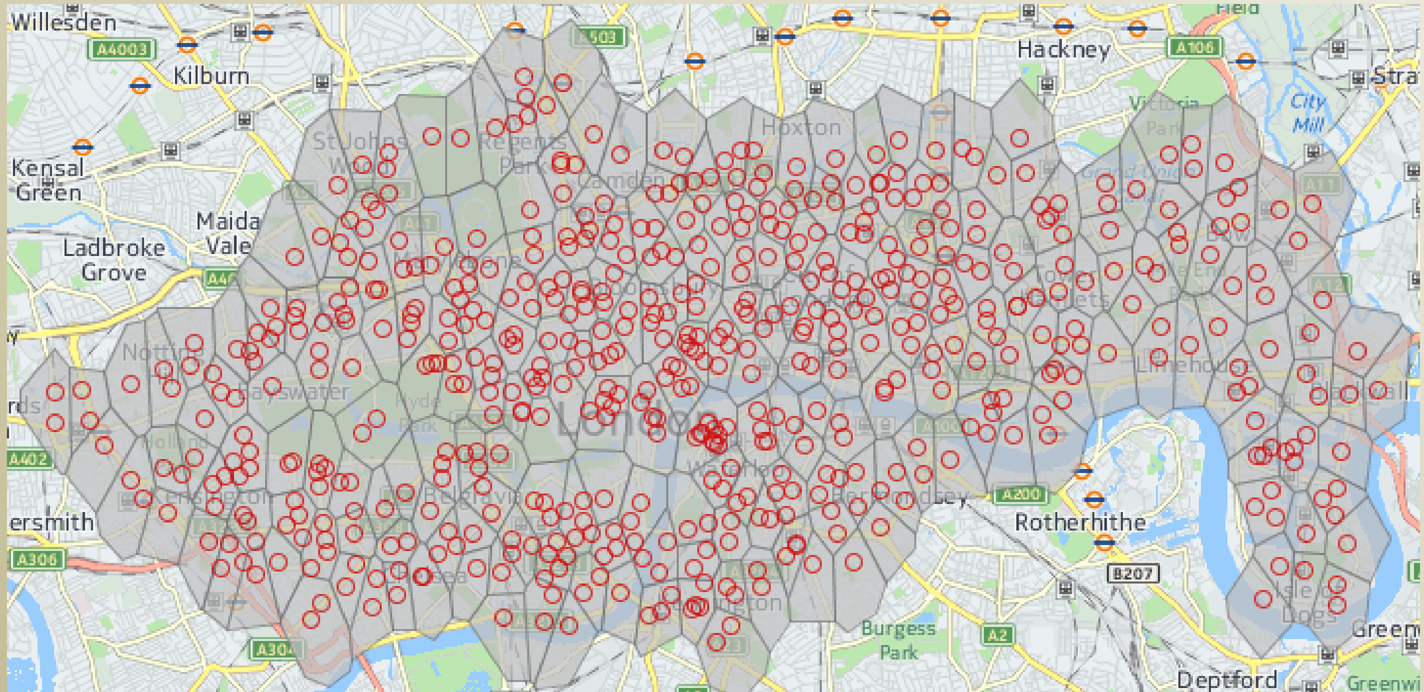


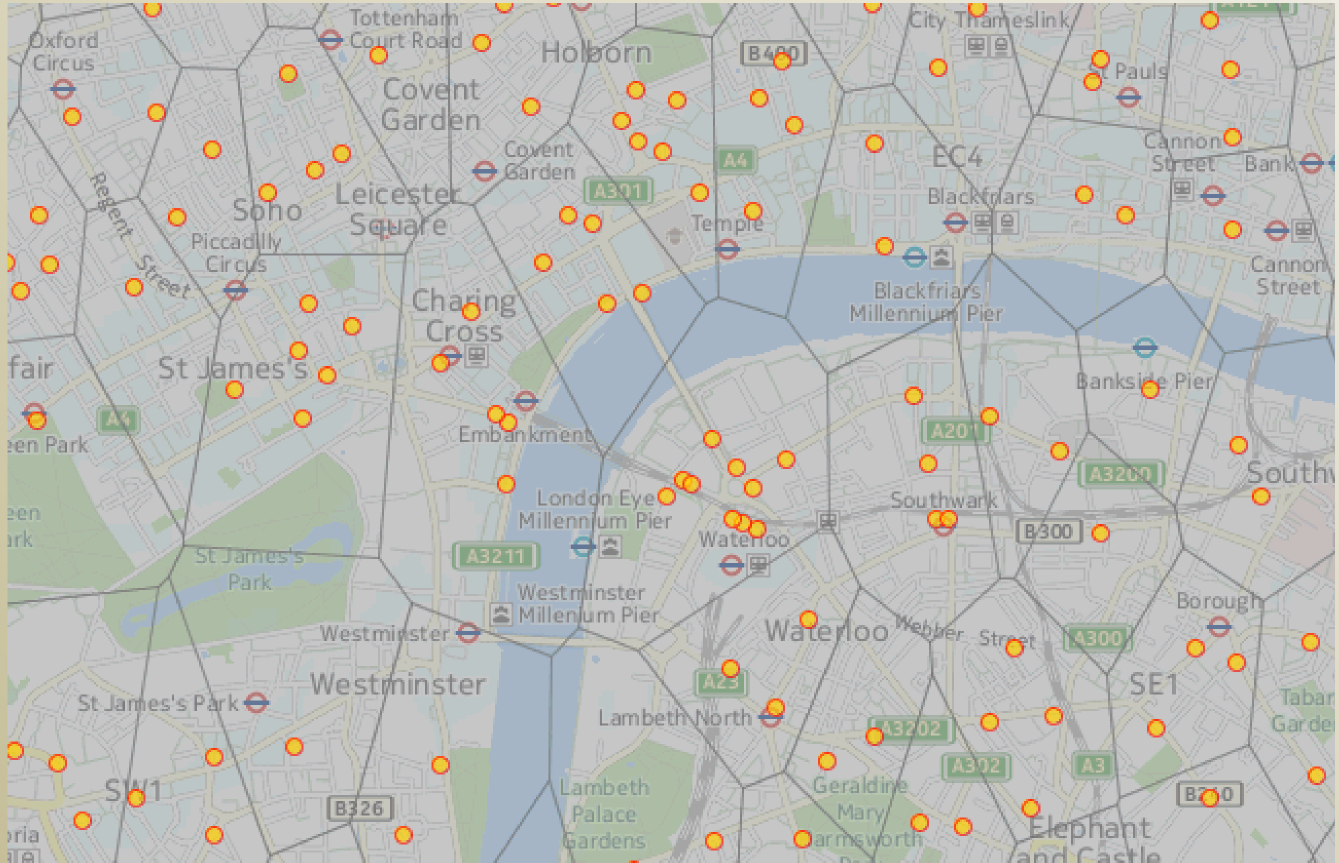
How do we obtain a tessellation?

Step 2: we use the centres of the point clusters (i.e., the centres of the circumcircles of the clusters) as the generating seeds for the Voronoi tessellation.

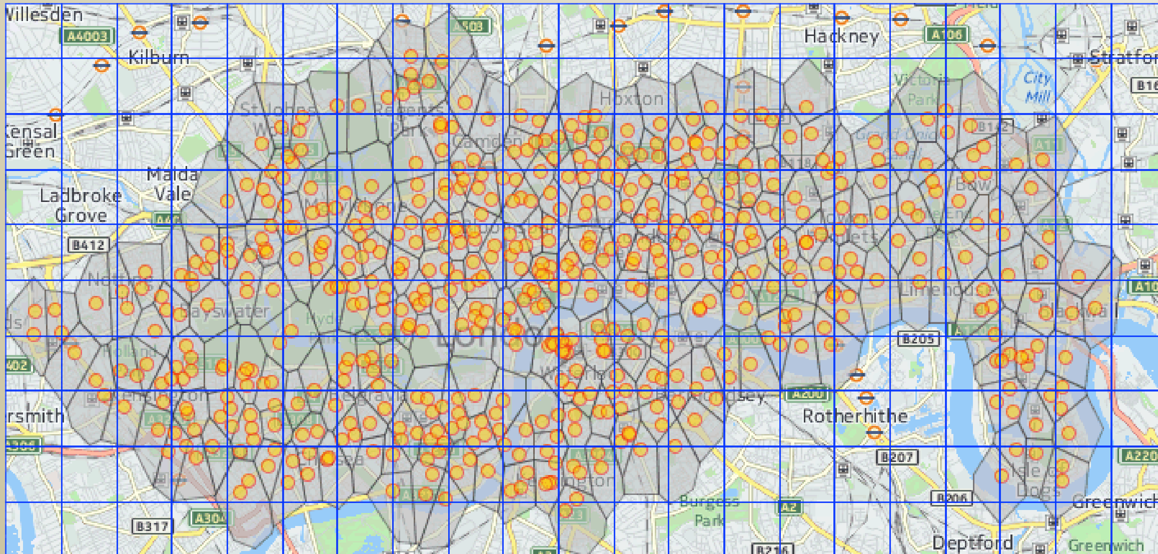


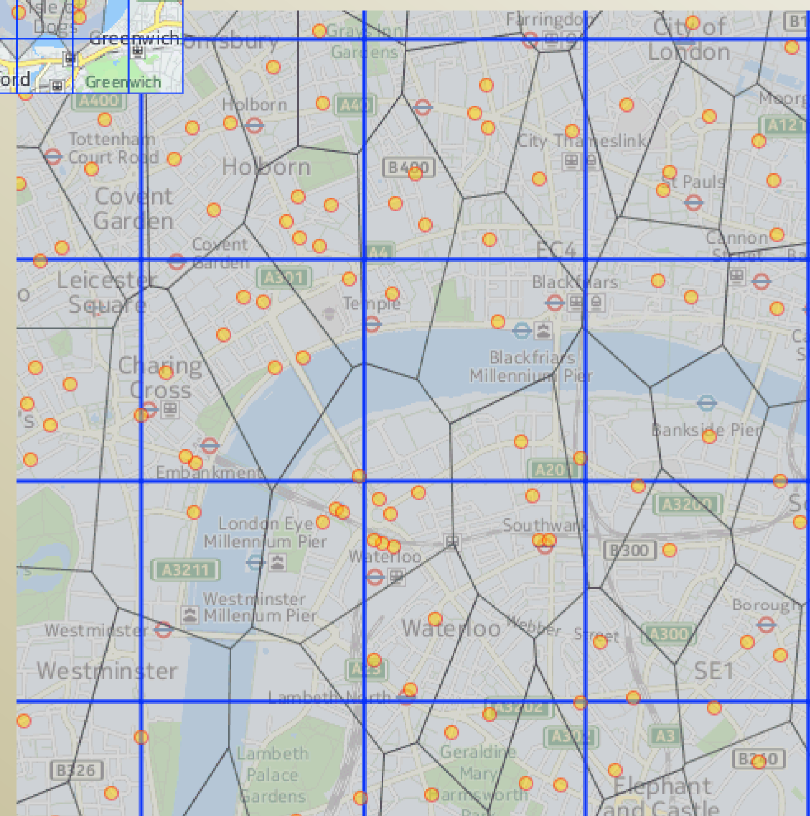
The empty cells (not containing any stations) can be removed.





The resulting irregular grid is built so that near stations tend to be included in the same cell. This allows trips with close origins or destinations to be counted together, e.g., trips from or to the area around Waterloo.

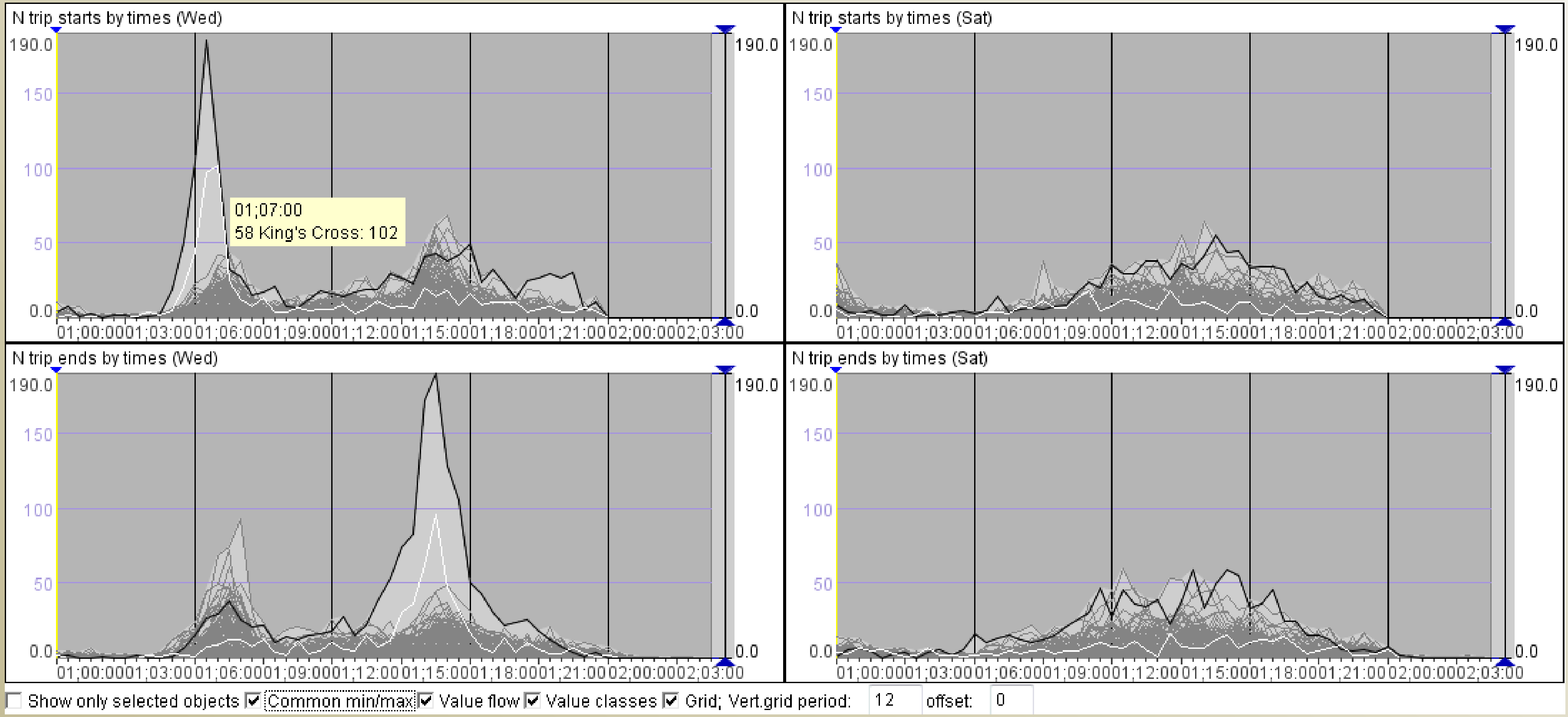




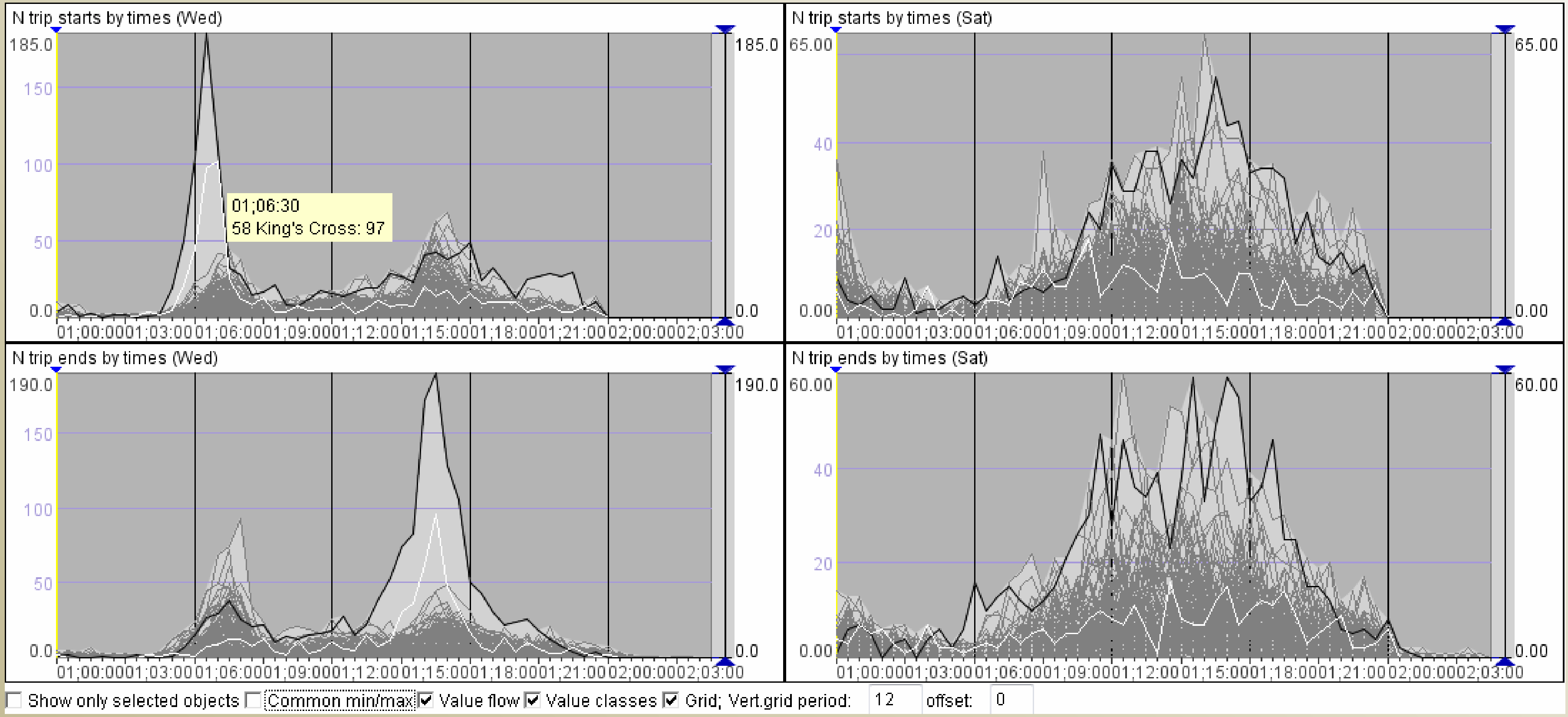
Compared to a regular grid, an irregular grid can give us a better idea about the spatial distribution of objects (in our example, trip origins and destinations).

Time division possibilities

- Treat the time as a directed <u>line</u>.
  - Divide the range from the earliest to the latest time value in the data into consecutive intervals.
  - To aggregate: for each time value in the data, find the containing interval.
- Treat the time as a <u>cycle</u>.
  - Choose a relevant cycle for the data: daily, weekly, annual, domain-specific, e.g., production cycle
  - Divide the chosen <u>cycle</u> into intervals.
  - To aggregate: for each time value in the data, determine its position in the cycle and find the containing interval of this position.
    - I.e., the absolute time value is transformed to relative w.r.t. the time cycle.



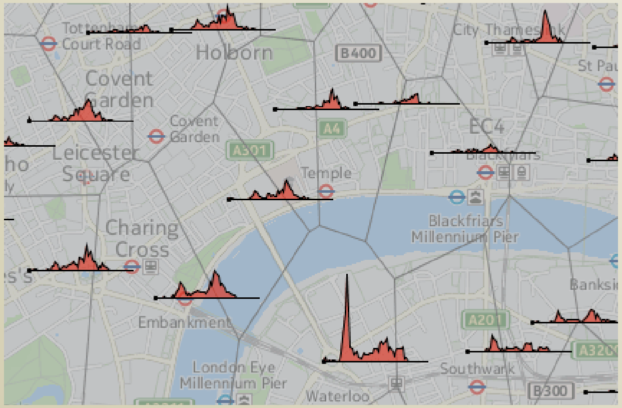
Example: for comparing the London bike trip data from Wednesday and Saturday, we have transformed the time values to the positions in the daily cycle and aggregated the data by 30-minutes intervals within the cycle. The time graphs show the counts of the trip starts (upper row) and ends (lower row) for the Wednesday 25/07/2012 (left) and Saturday 28/07/2012 (right). Highlighted in black are the curves for the Waterloo area and in white for the King's cross area.

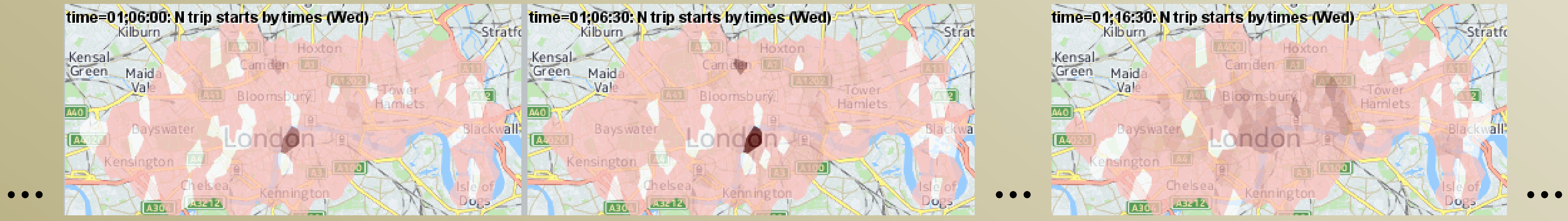


Here, each time graph has its own vertical scale.

Aggregation result: spatial time series

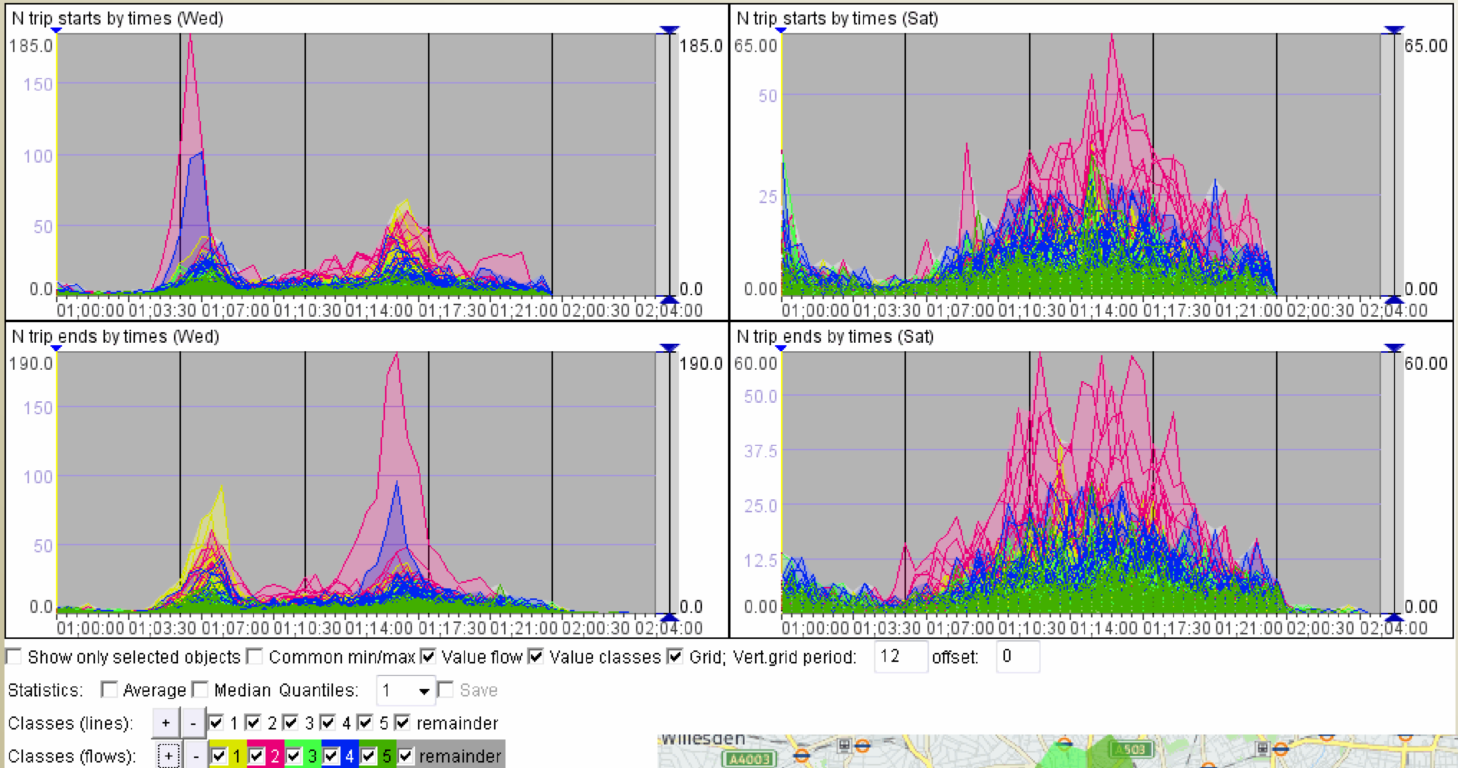
- Data with two referrers $S \times T \rightarrow A$, or $A(s,t)$, where $s \in S$, $t \in T$
  - S is the set of space compartments and T is the set of time intervals
- Can be viewed in two complementary ways (like *projections*):
  - as a set of time series of attribute values associated with the spatial compartments
  - as a set of *spatial situations* associated with the time steps
    - A *spatial situation* is a distribution of attribute values over the space in some time step



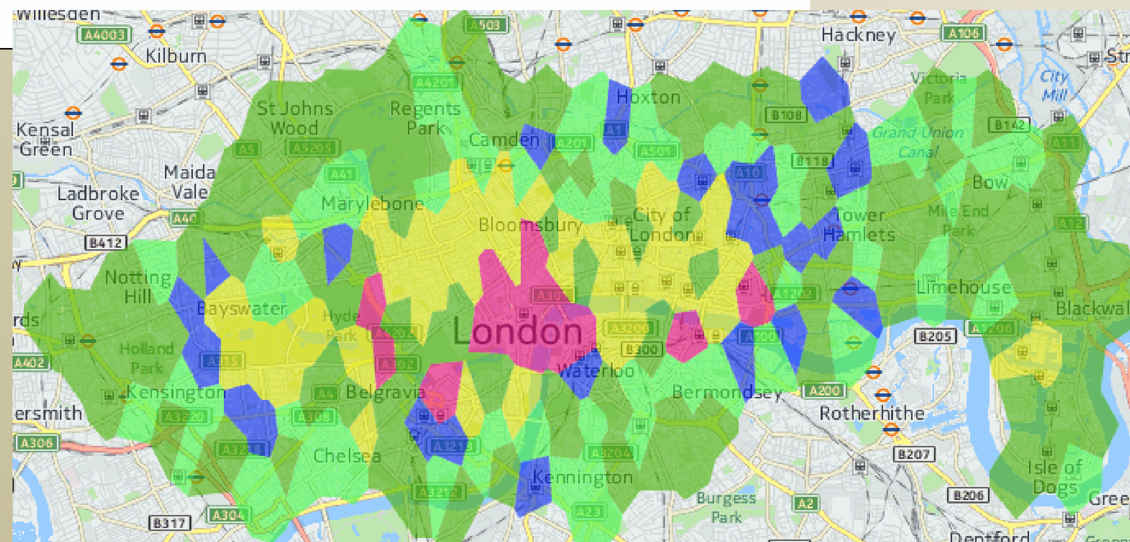


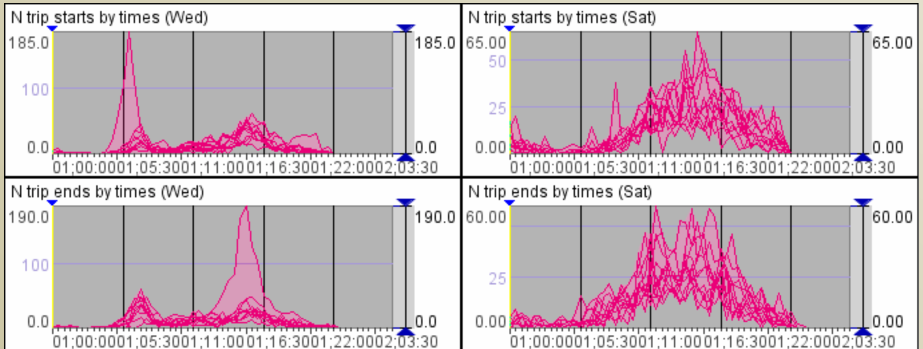
Analysis of spatial time series *(a reminder)*

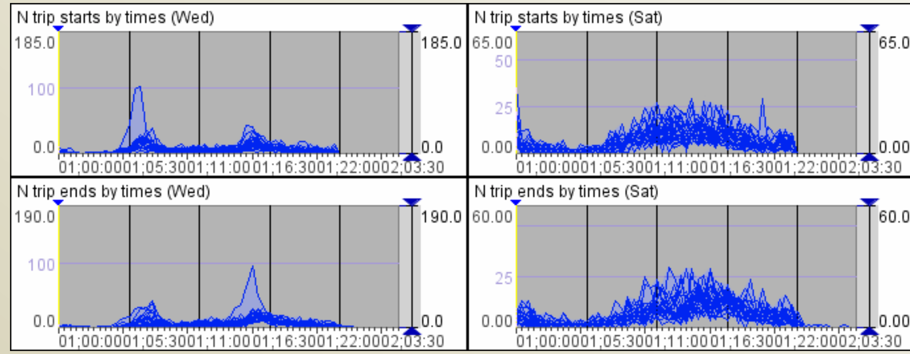
- Analysis tasks address two aspects of the overall behaviour:
  - Spatial distribution of the local temporal variations of the attribute values in different compartments
  - Temporal variation of the overall spatial distribution of the attribute values in different time moments
- Supporting visual analytics techniques *(considered in the previous lecture)* include two-way partition-based clustering
  - Grouping of places (compartments) by similarity of the local time series
    - followed by visual exploration of the distribution of the group members over the space
  - Grouping of time steps by similarity of the spatial situations
    - followed by visual exploration of the distribution of the group members over time

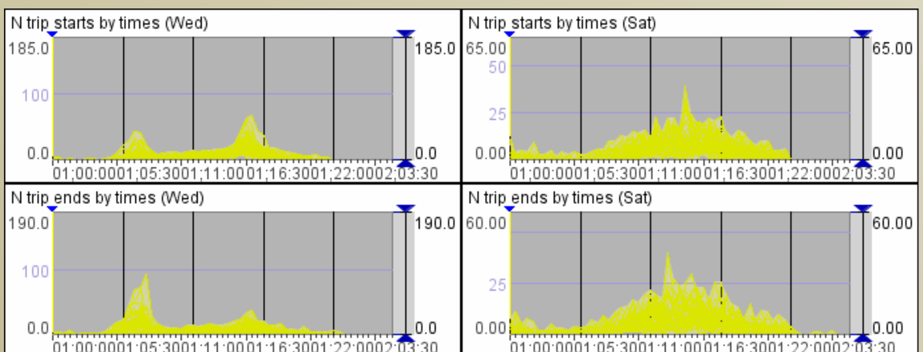


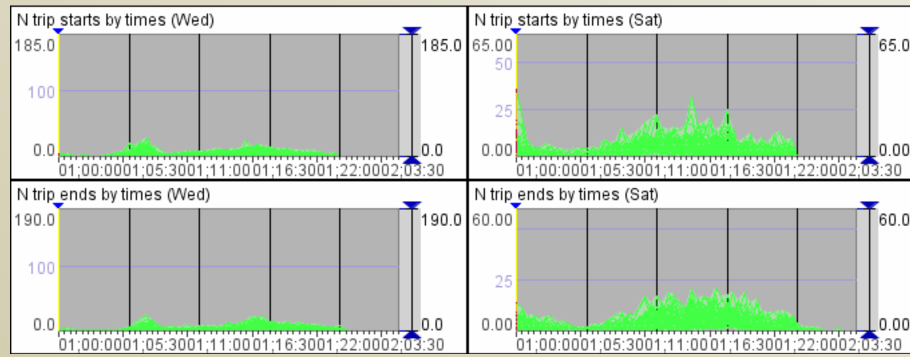
Example 1: the spatial compartments (grid cells) have been clustered by the similarity of the time series of the counts on the trip starts and ends on Wednesday and Saturday.

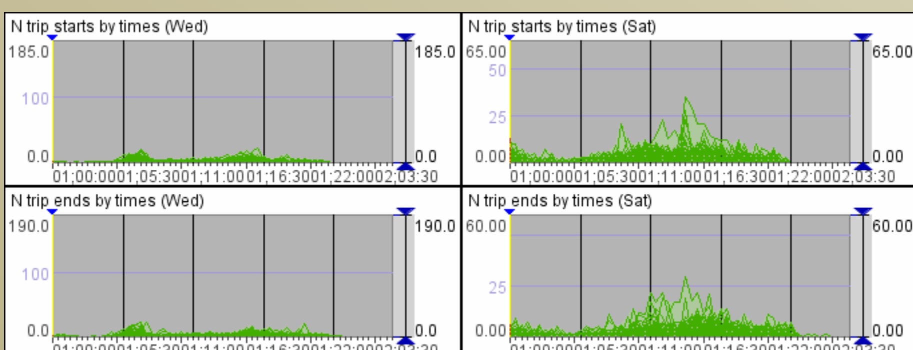


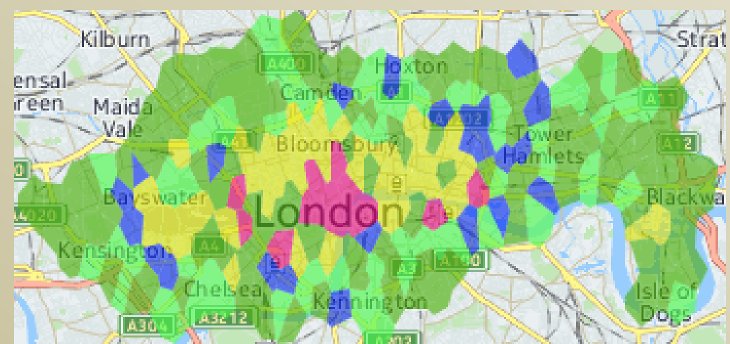

N trip starts by times (Wed)
185.0
100
0.0
N trip starts by times (Sat)
65.00
50
25
0.00
N trip ends by times (Wed)
190.0
100
0.0
N trip ends by times (Sat)
60.00
25
0.00
01;00:00 01;05:30 01;11:00 01;16:30 01;22:00 02;03:30

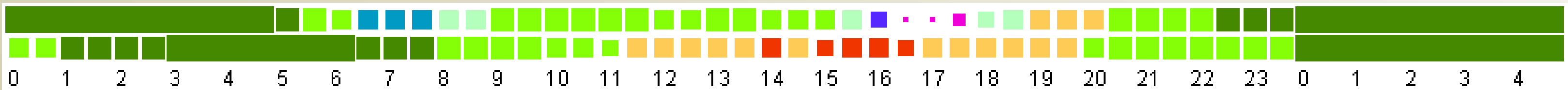

N trip starts by times (Wed)
185.0
100
0.0
N trip starts by times (Sat)
65.00
50
25
0.00
N trip ends by times (Wed)
190.0
100
0.0
N trip ends by times (Sat)
60.00
25
0.00
01;00:00 01;05:30 01;11:00 01;16:30 01;22:00 02;03:30


N trip starts by times (Wed)
185.0
100
0.0
N trip starts by times (Sat)
65.00
50
25
0.00
N trip ends by times (Wed)
190.0
100
0.0
N trip ends by times (Sat)
60.00
25
0.00
01;00:00 01;05:30 01;11:00 01;16:30 01;22:00 02;03:30

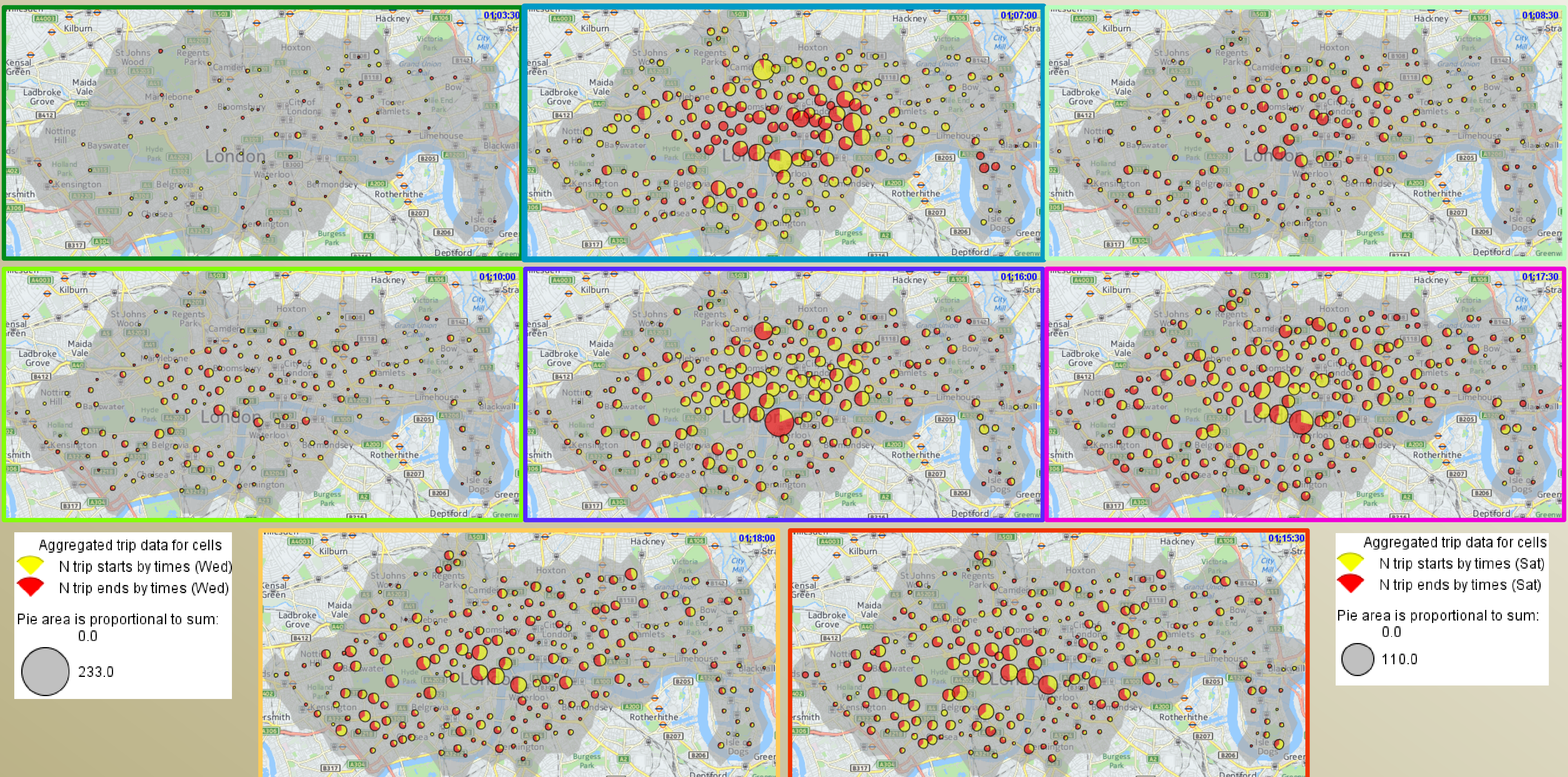

N trip starts by times (Wed)
185.0
100
0.0
N trip starts by times (Sat)
65.00
50
25
0.00
N trip ends by times (Wed)
190.0
100
0.0
N trip ends by times (Sat)
60.00
25
0.00
01;00:00 01;05:30 01;11:00 01;16:30 01;22:00 02;03:30


N trip starts by times (Wed)
185.0
100
0.0
N trip starts by times (Sat)
65.00
50
25
0.00
N trip ends by times (Wed)
190.0
100
0.0
N trip ends by times (Sat)
60.00
25
0.00
01;00:00 01;05:30 01;11:00 01;16:30 01;22:00 02;03:30


Kilburn
Maida Vale
Hoxton
Camden
Bloomsbury
Tower Hamlets
Bayswater
London
Kensington
Chelsea
Kennington
Isle of Dogs



Example 2: clustering of the 30-minutes time intervals by the similarity of the spatial distributions of the trip starts and ends. Upper row of the time arranger: Wednesday, lower row: Saturday.



Questions?

Analysis of aggregated OD movement data referring to places (space compartments)

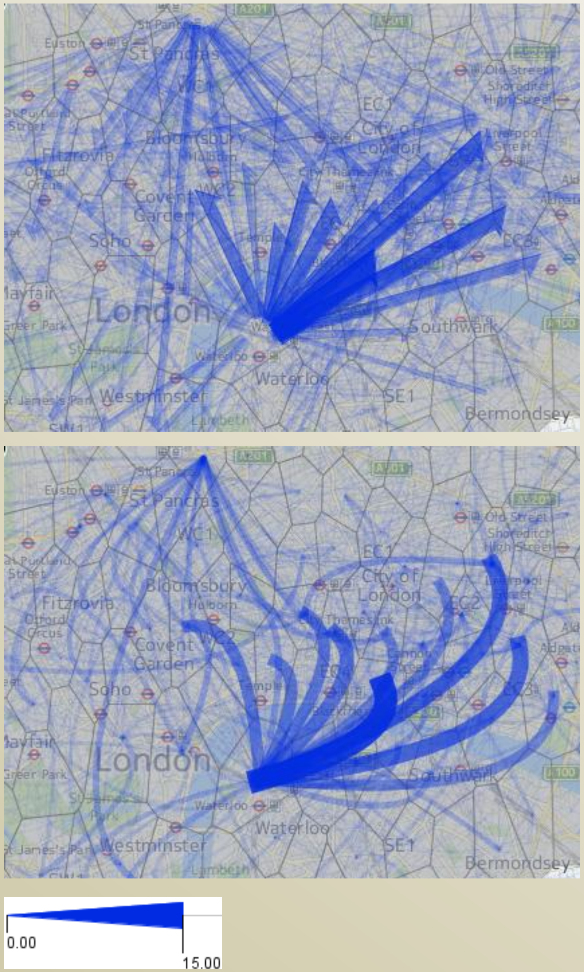
Trip aggregation by OD pairs

- The ST aggregation considered before loses information about the connections between the trip origins and destinations.
- To preserve this information, we need to aggregate data differently: by pairs of compartments and time intervals.
  - From the set of the spatial compartments S, we create the set of all possible pairs of compartments, i.e., the Cartesian product $S \times S$
  - As previously, we divide the time (line or cycle) into a set of intervals T*
  - To aggregate the trip data (OD moves):
    - For each pair (s_o, s_d) and each time interval t, find and summarize all trips that started in s_o and ended in s_d within the interval t*.

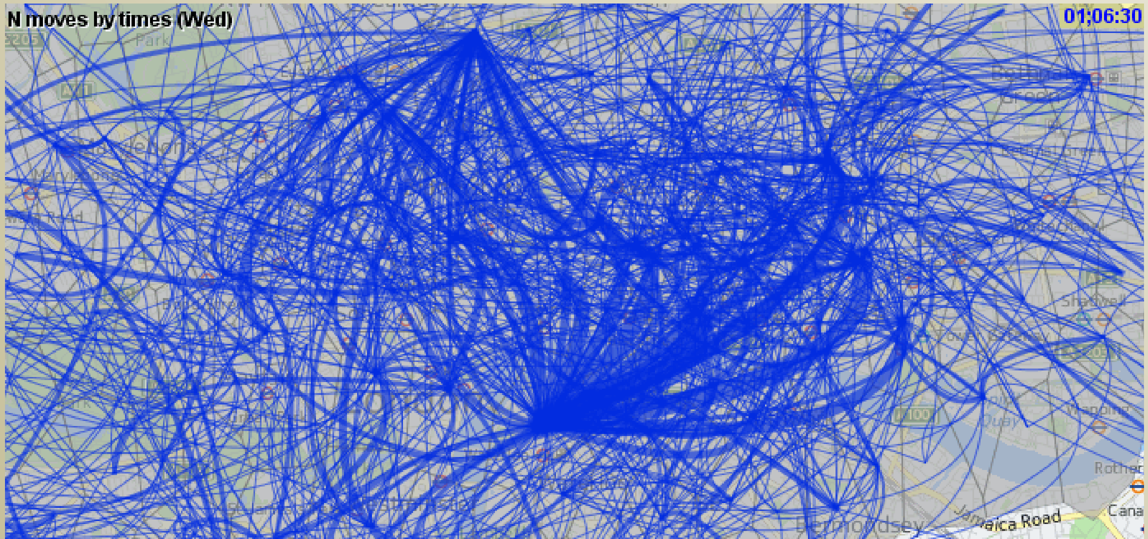
* It would also be appropriate to create the set of all possible pairs of time intervals (t_i, t_j), where $t_i \le t_j$, and aggregate the data by the combinations (s_o, s_d, t_i, t_j); however, the result would be too complex for the analysis. For simplification, trips can be assigned to time intervals based on the times of their ends or starts.

Result of trip aggregation by OD pairs

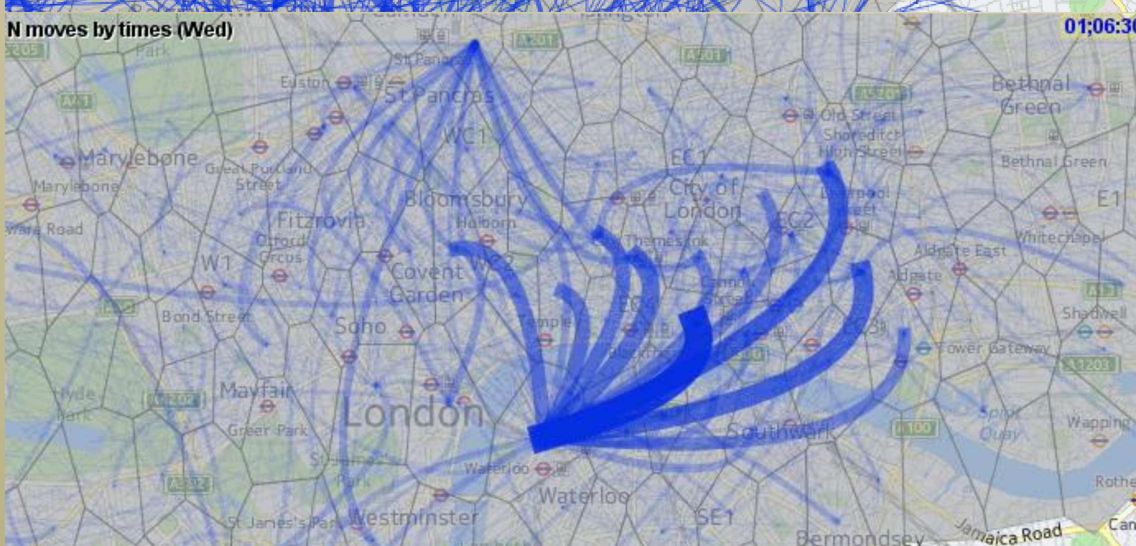
- Aggregated trip data have 3 referrers *(it would be 4 without the temporal simplification* : $S \times S \times T$
  - Attributes $A(s_o, s_d, t)$: trip counts + statistical summaries of the trip characteristics (e.g., durations)
- It is very hard to visualise and analyse data with 3 referrers
- Approach: treat pairs (s_o, s_d) as a special type of geographic objects, called *spatial links*, or *flows*
  - Treat the aggregate attributes as usual spatial time series referring to these spatial objects
- Visual representation of flows on a map: flow symbols; widths and/or colours may represent values of thematic attributes.



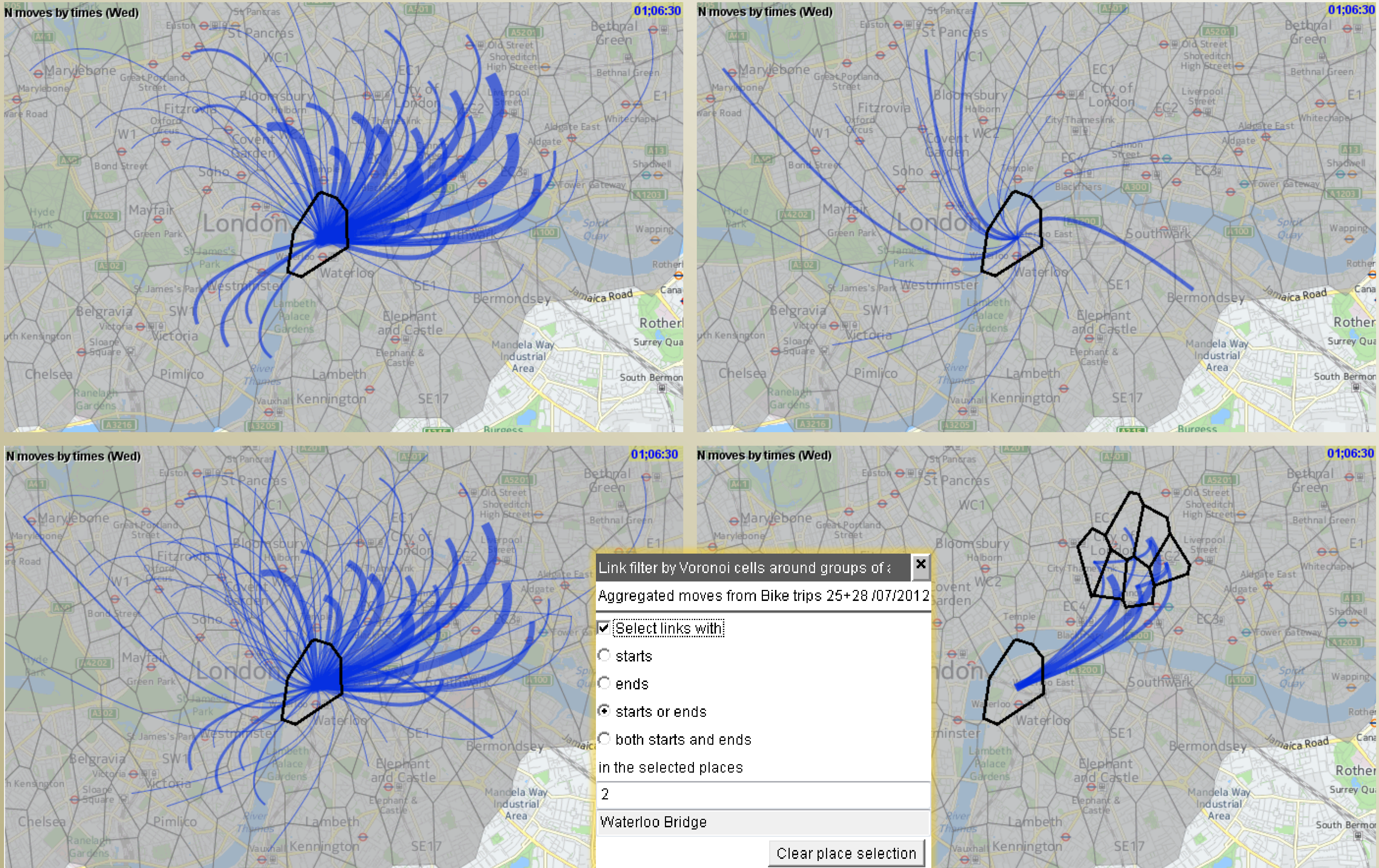
Problems of flow visualisation



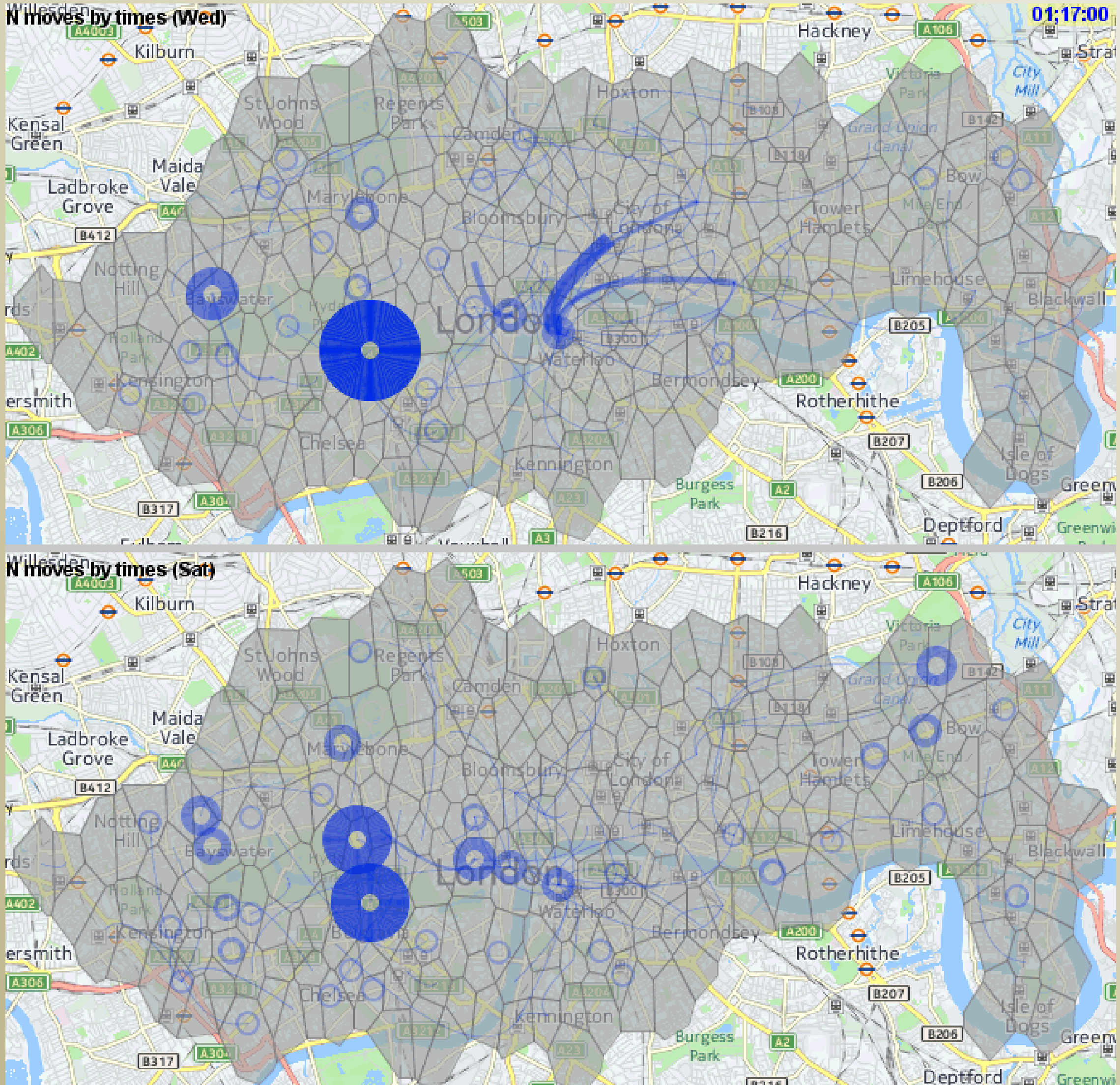
Representing flows on a map typically results in extremely cluttered images due to numerous intersections of the flow symbols.



The problem may be reduced by varying the transparency of the flow symbols depending on the attribute values that are represented.

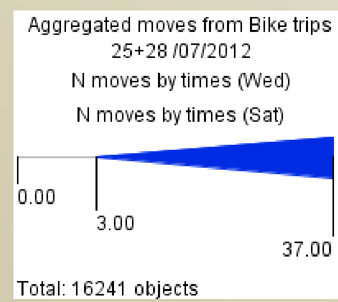


Another approach: interactive filtering by the origins and/or destinations.



Flows with coinciding origin and destination places may be represented by rings with the thickness proportional to values of a numeric attribute, such as the count of the moves.

For decreasing the display clutter, flows with small attribute values can be hidden by means of interactive focusing.



Spatial link-referenced time series

- We transformed the complex data structure $S \times S \times T \rightarrow A$ into a simpler data structure $L \times T \rightarrow A$
  - L is the set of *spatial links* defined as pairs (s_o, s_d), $s_o \in S$, $s_d \in S$.
- Hence, we obtained "normal" space-referenced time series, in which attribute values refer to spatial objects (specifically, the spatial links) and time steps.

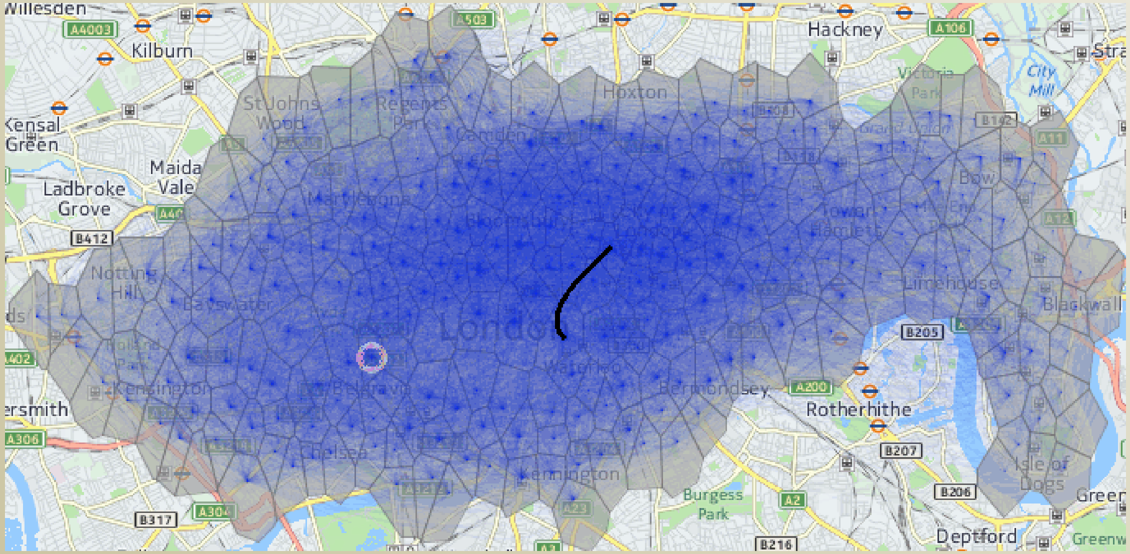
$\Rightarrow$ We can apply the same visual analytics techniques as for usual time series

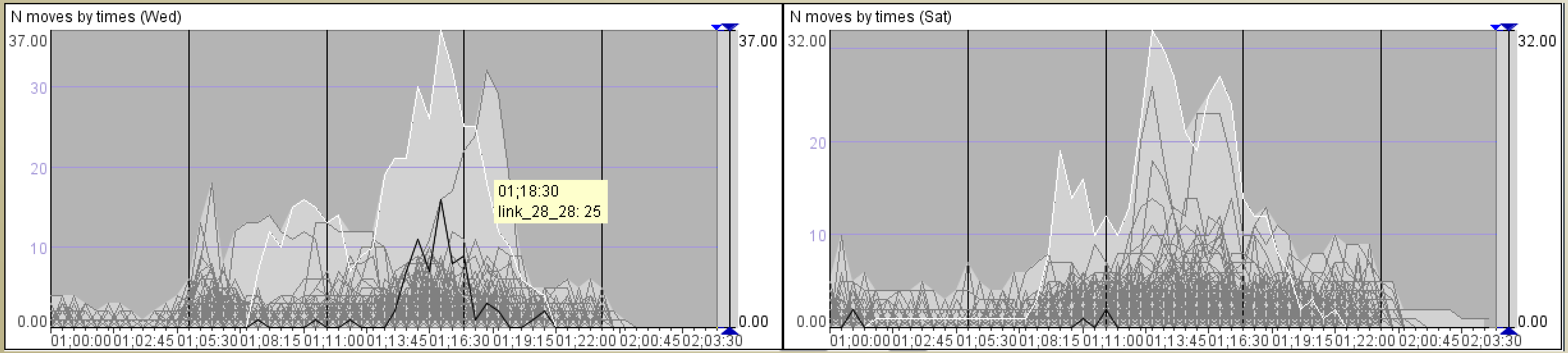
- … but we need to deal with the difficulties in the visual representation of the spatial links in a map

  ☹ This is a visualisation problem that is not completely solved yet.

View (projection) 1

Set of time series associated with the links

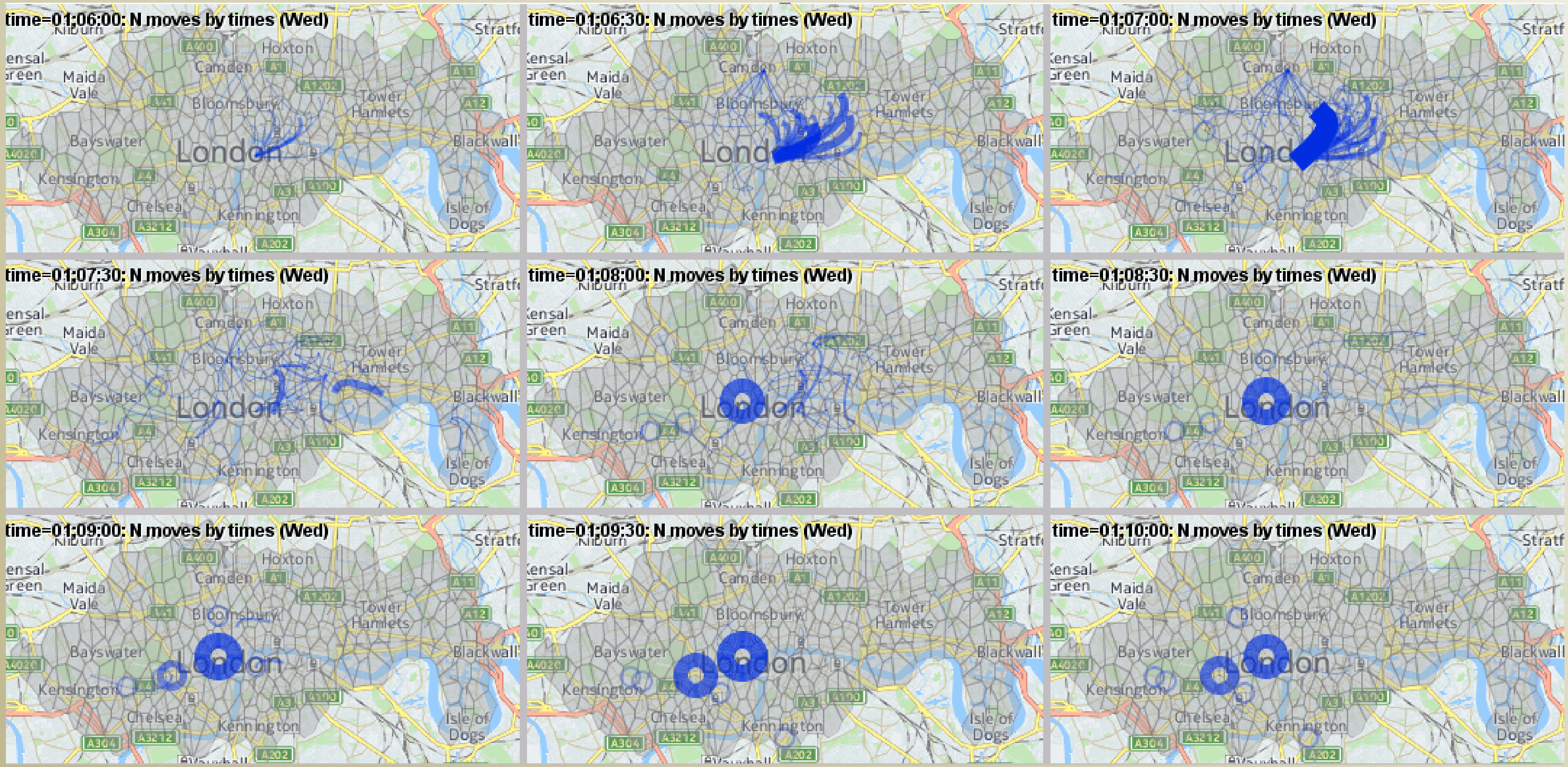




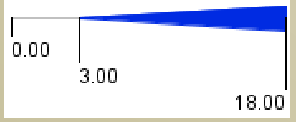
View (projection) 2

Set of spatial situations associated with the time steps

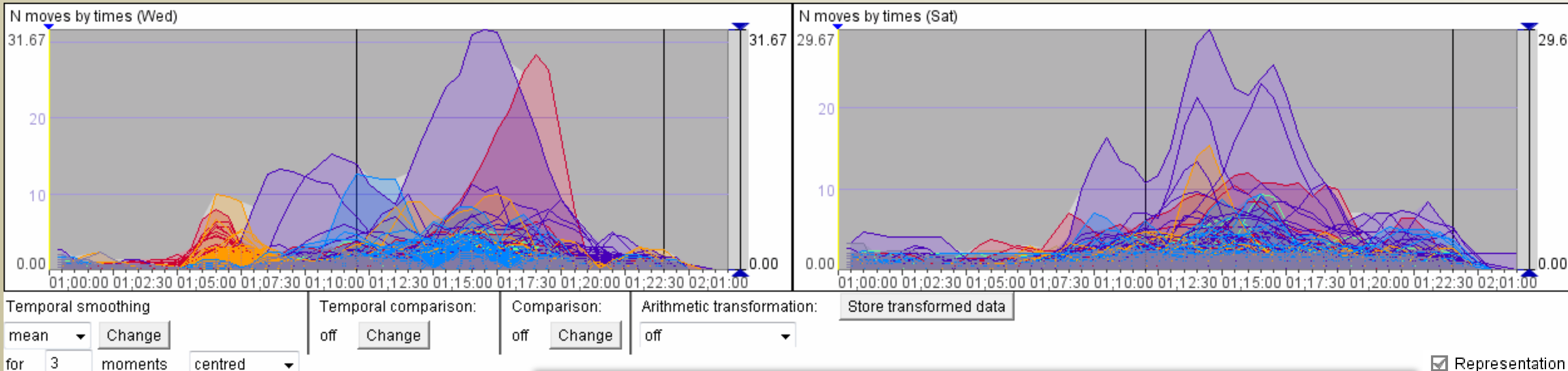
…



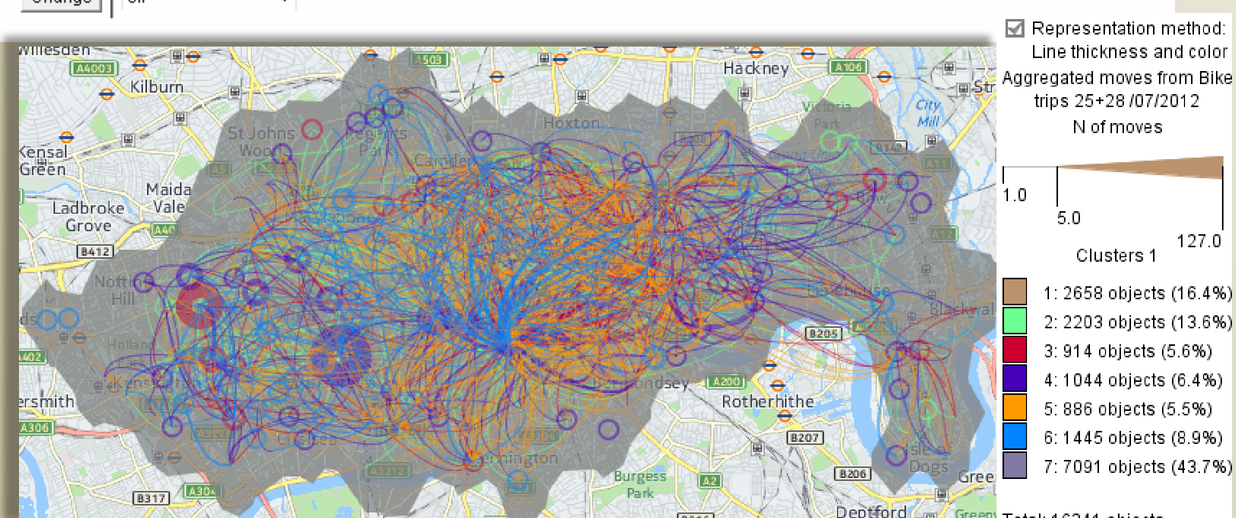
…



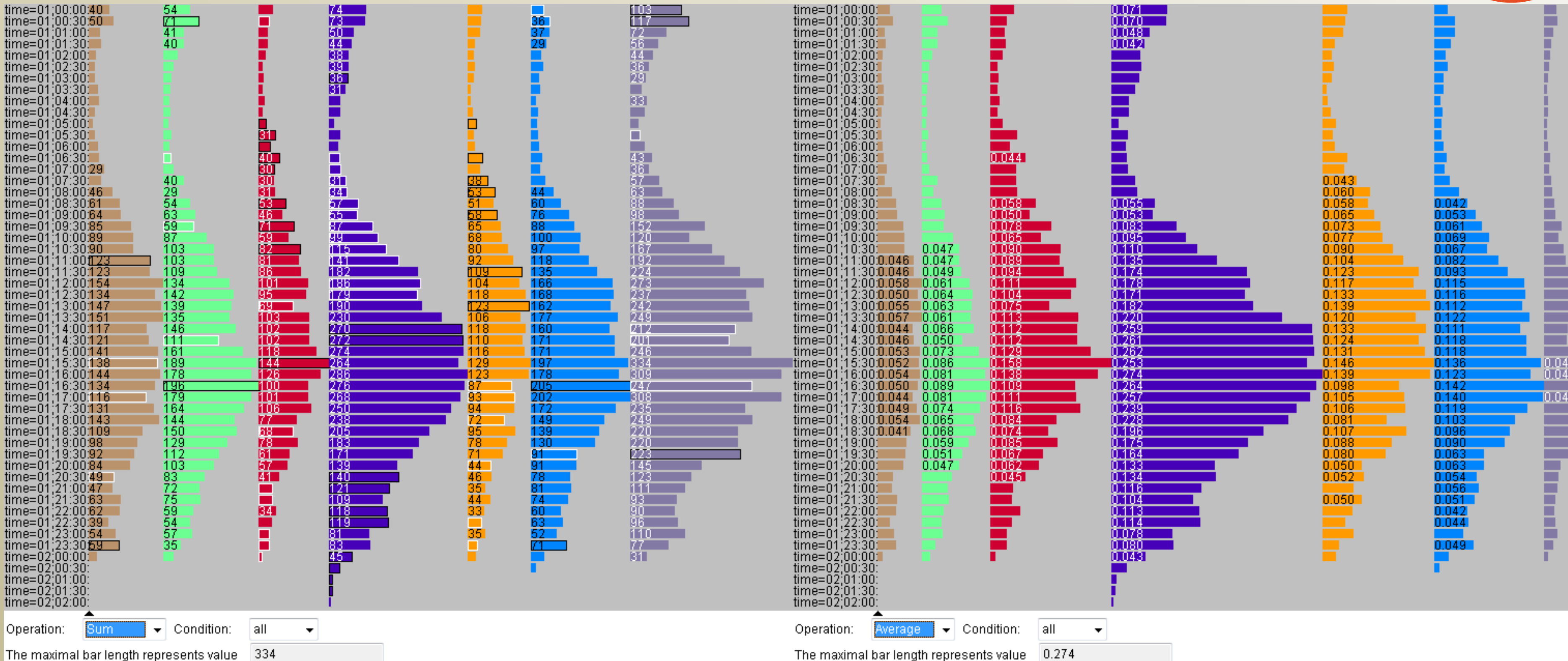
Investigating the distribution of the temporal behaviours over the set of links

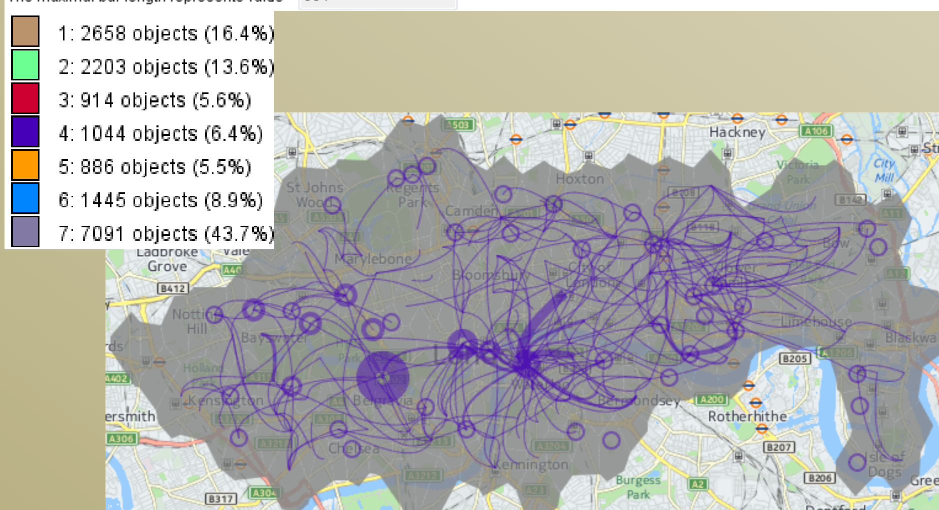


By applying partition-based clustering to the time series associated with the links, we divide the links according to the temporal profiles of their use over the two days.

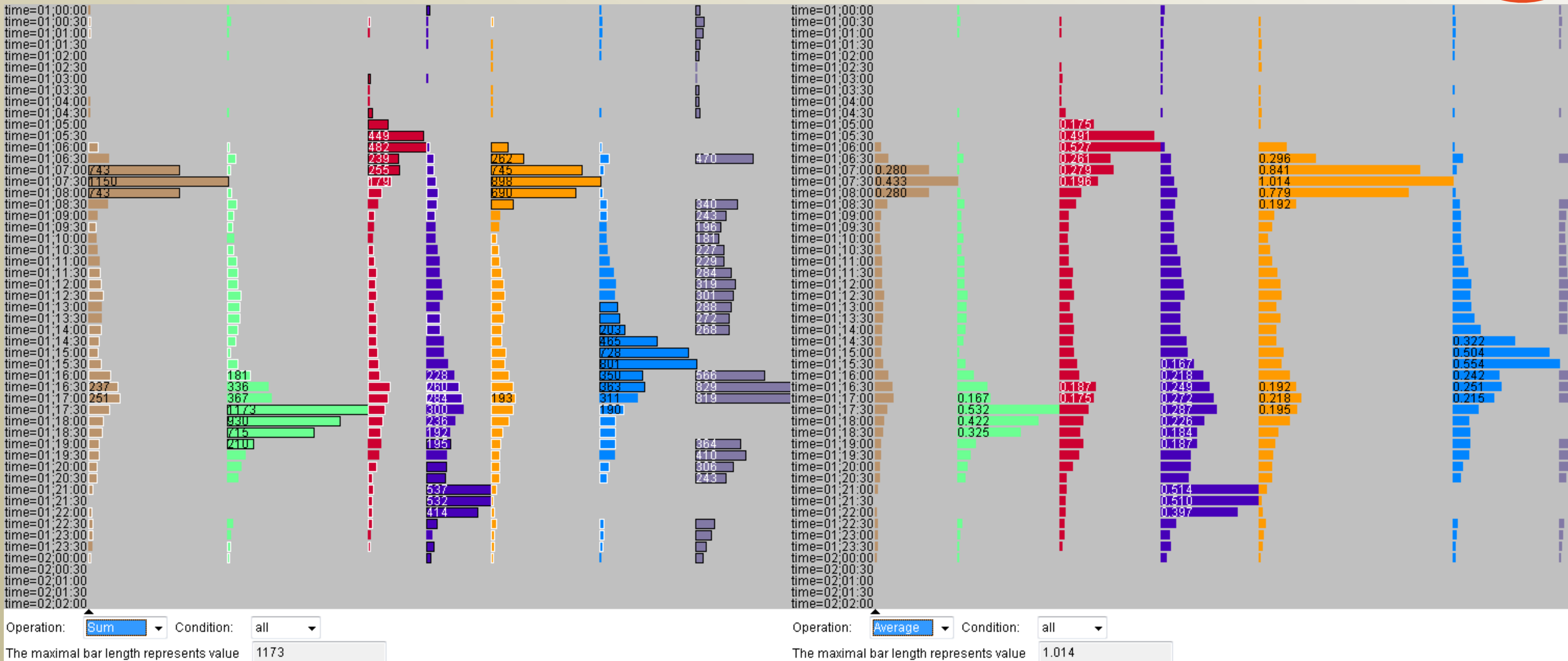


The map shows only the links that were used at least 5 times during the two days. The widths of the flow symbols are proportional to the total counts of the moves.





On Saturday, all link clusters have similar patterns of the temporal behaviour of the move counts. The clusters differ only in the magnitudes of the counts. Cluster 4 (violet) includes the links that were the most actively used on Saturday.

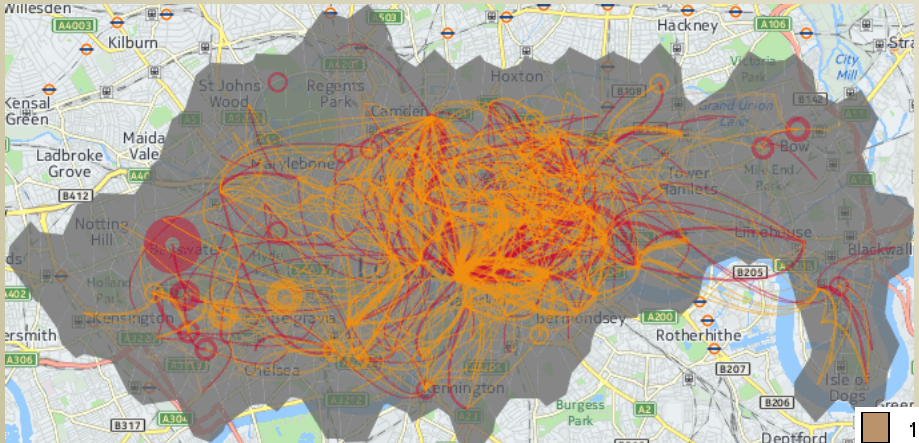


1: 2658 objects (16.4%)
2: 2203 objects (13.6%)
3: 914 objects (5.6%)
4: 1044 objects (6.4%)
5: 886 objects (5.5%)
6: 1445 objects (8.9%)
7: 7091 objects (43.7%)

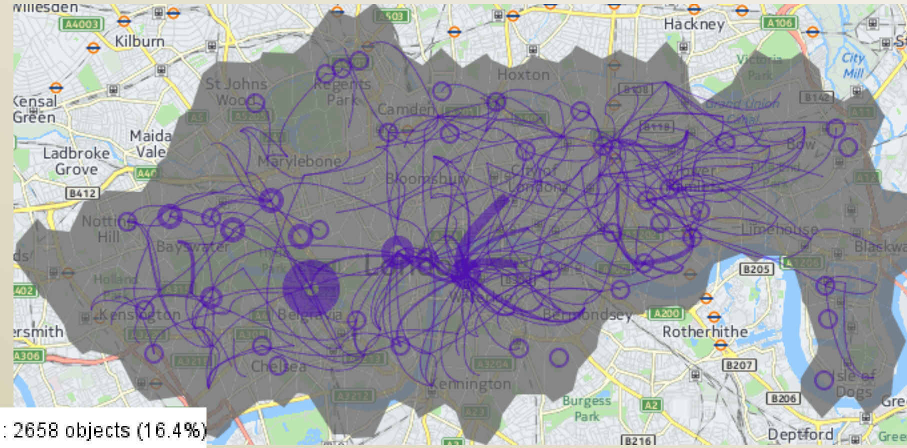
On Wednesday, the temporal behaviours of the move counts on the links substantially differed. The links of clusters 1, 3, and 5 were predominantly used in the morning and the links of clusters 2, 4, and 7 in the afternoon. These clusters may be related to the trips of the bike users to and from their places of work or study.

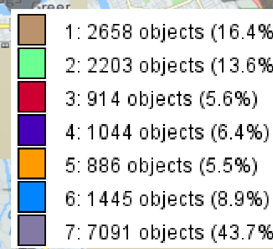
The links of cluster 4 were more intensively used in the evening. Since they were also intensively used on Saturday, we can conjecture that they may be related to leisure bike trips.

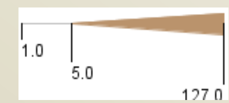
Wednesday morning link clusters



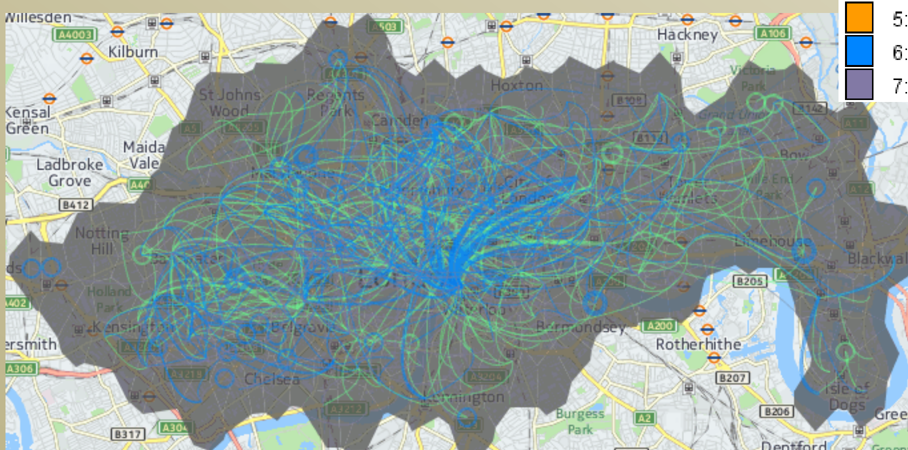
Leisure time link clusters

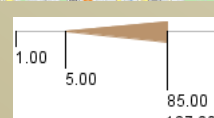






Wednesday afternoon link clusters

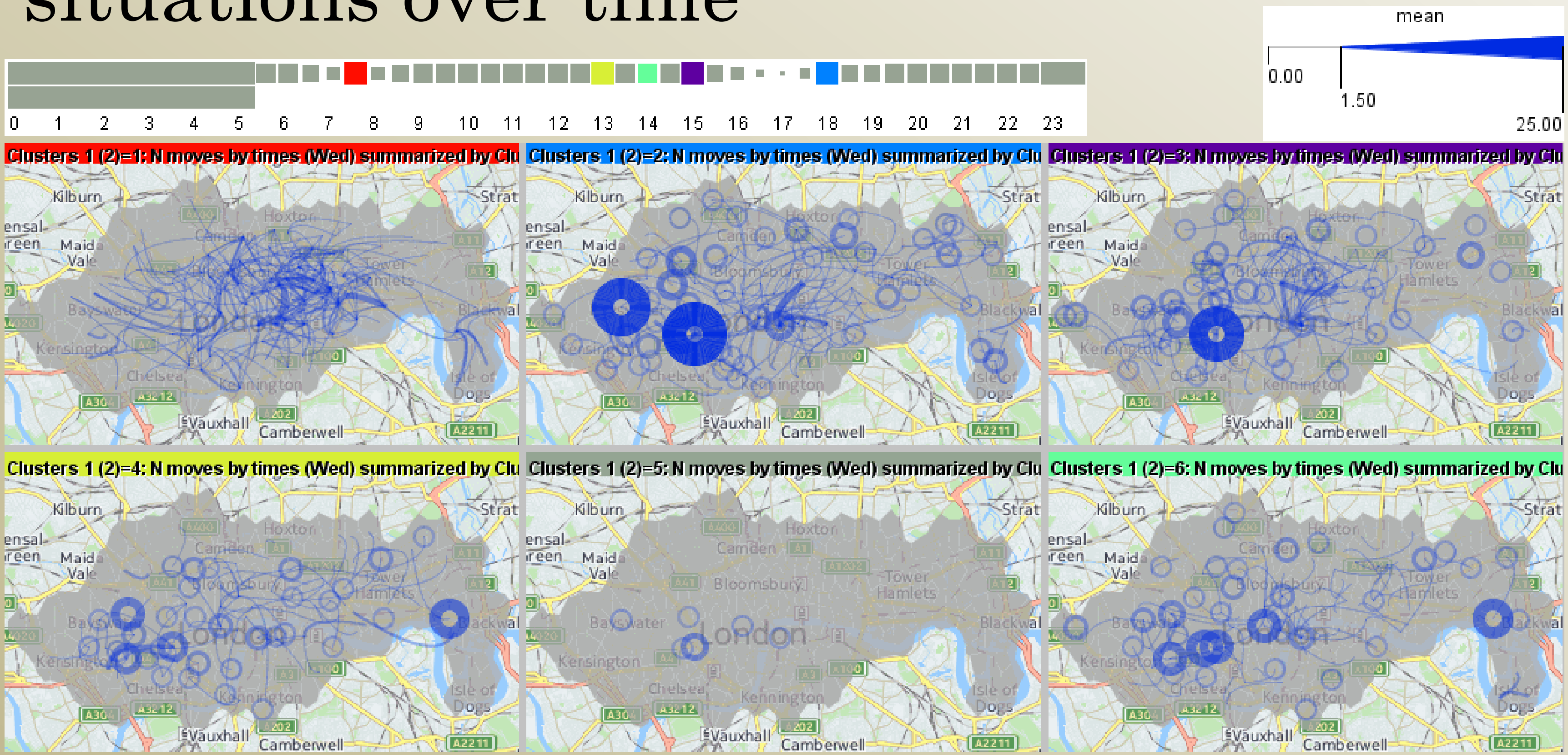




The interpretation of the clusters is not completely “clean”. Thus, it appears that clusters 3 and 5 may include “leisure” links, which might also be used by some people on Wednesday morning.

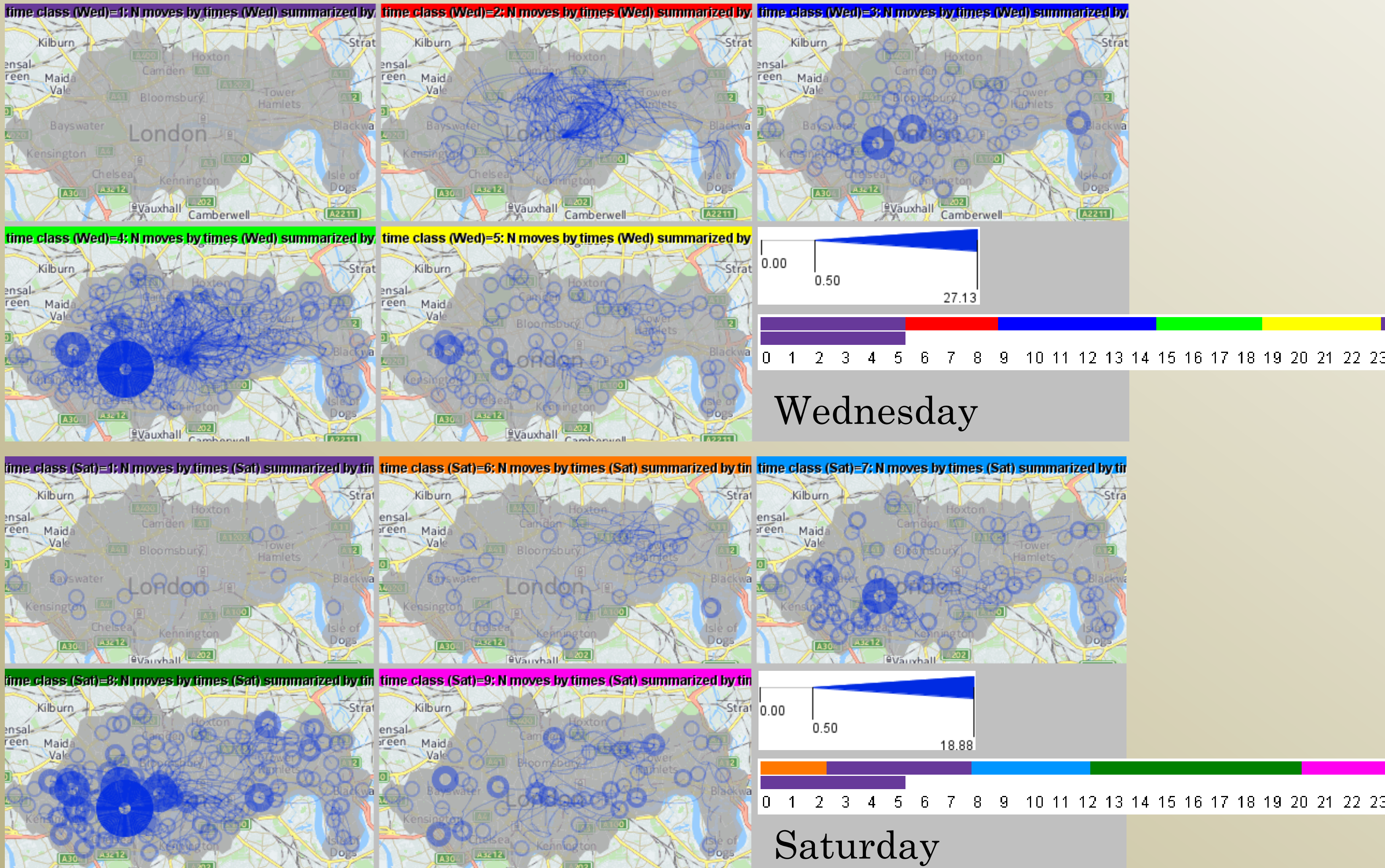
Even when clustering does not give us clear separation, it still reveals the major patterns and facilitates our understanding.

Investigating the variation of the spatial situations over time



Partition-based clustering has been applied to the combinations of the move count values for the links corresponding to the different time intervals of the Wednesday. Only the links with the total move counts ≥ 3 were used (4,897 of 16,241, or 30.2%). The clustering does not produce meaningful groups but just isolates certain intervals.

Interactive manual partitioning of the set of time steps based on background knowledge or previous results

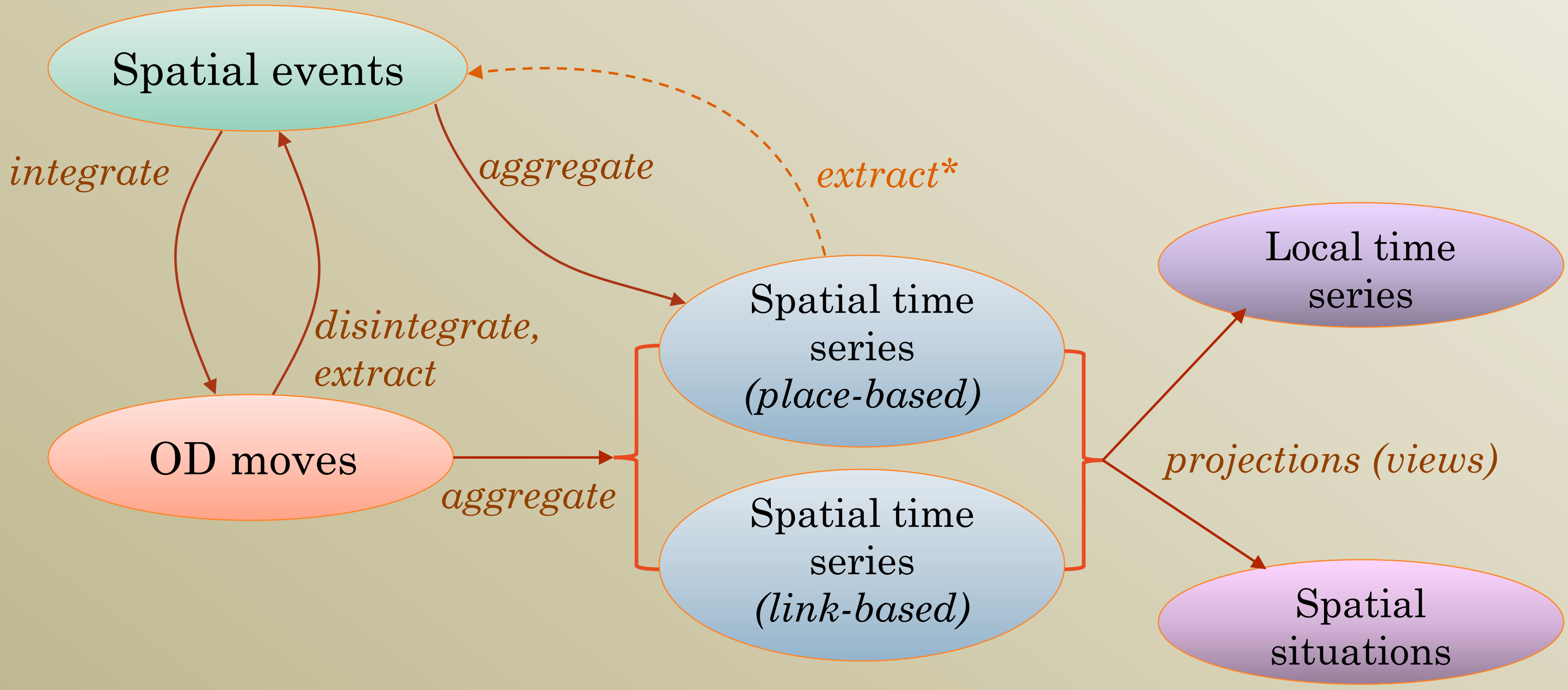


Questions?

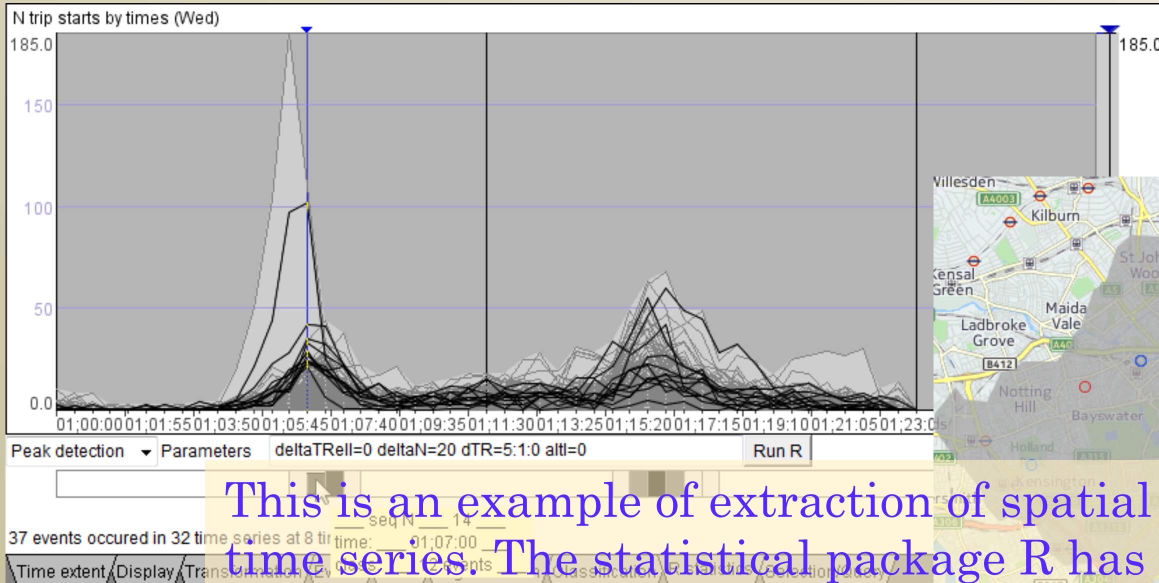
Aggregation of OD movement data by OD pairs and analysis of flows

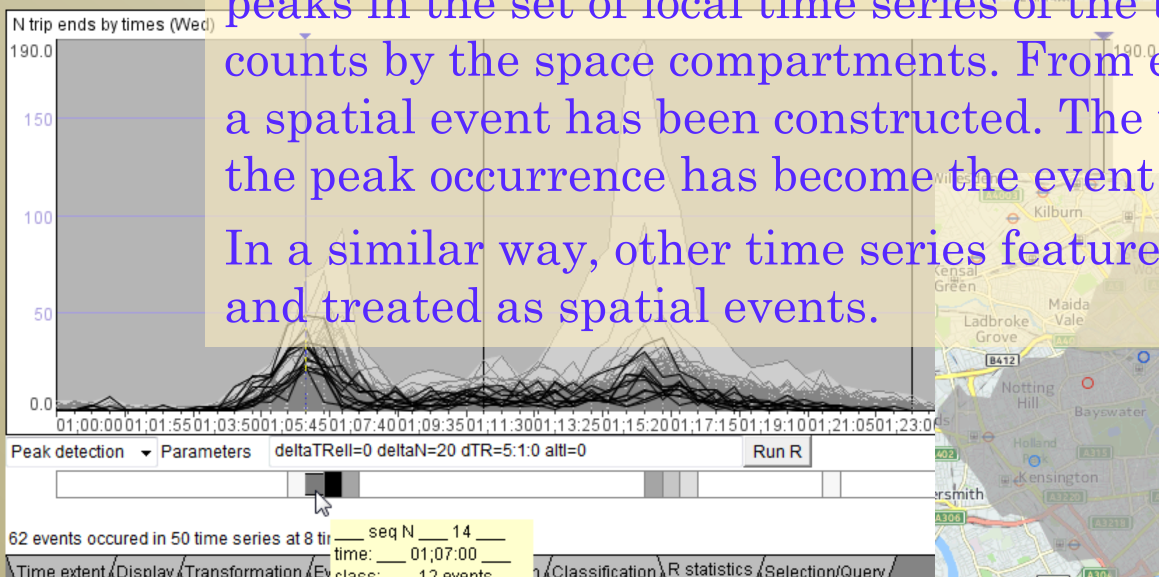
Types of spatio-temporal data

and transformations between them



* Not yet considered; shown in the following slides.

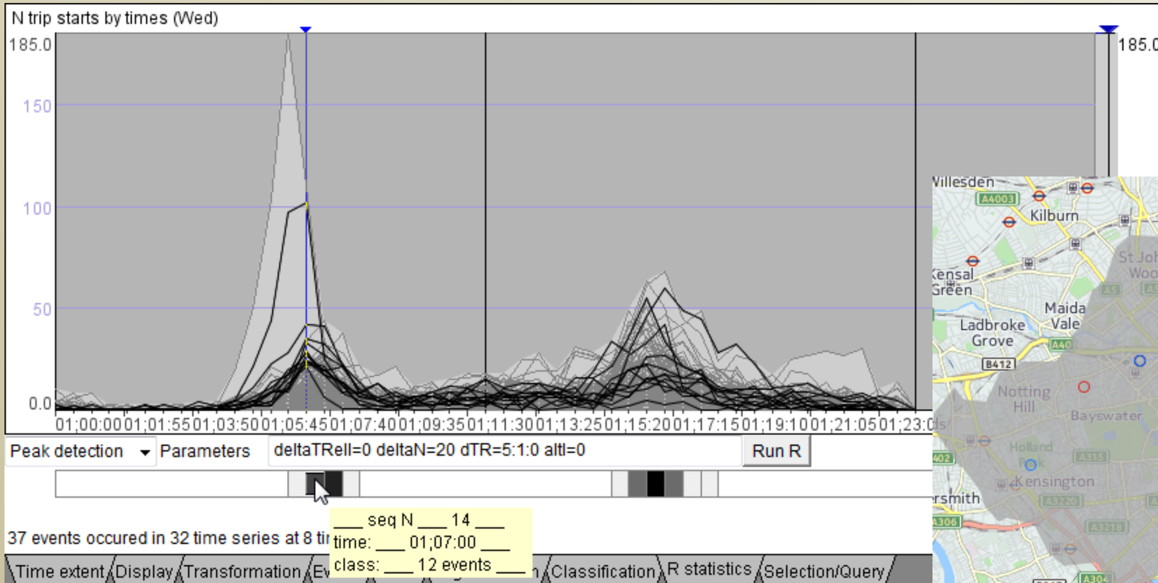


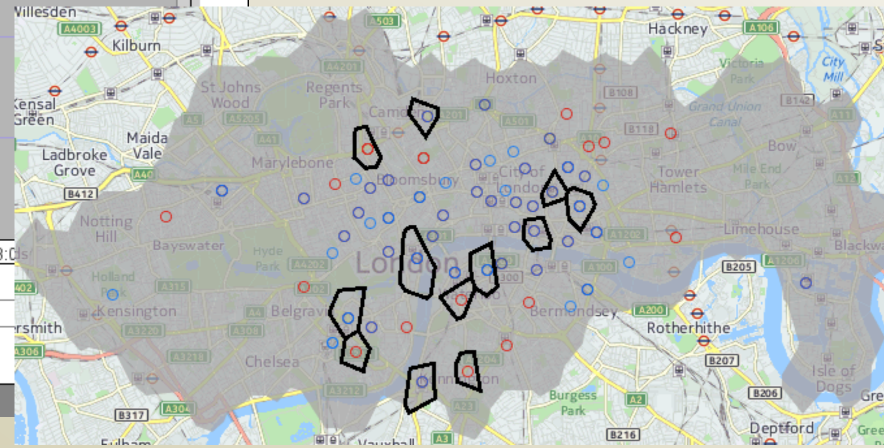


This is an example of extraction of spatial events from spatial time series. The statistical package R has been applied to find peaks in the set of local time series of the trip start and end counts by the space compartments. From each detected peak, a spatial event has been constructed. The time and place of the peak occurrence has become the event time and place.

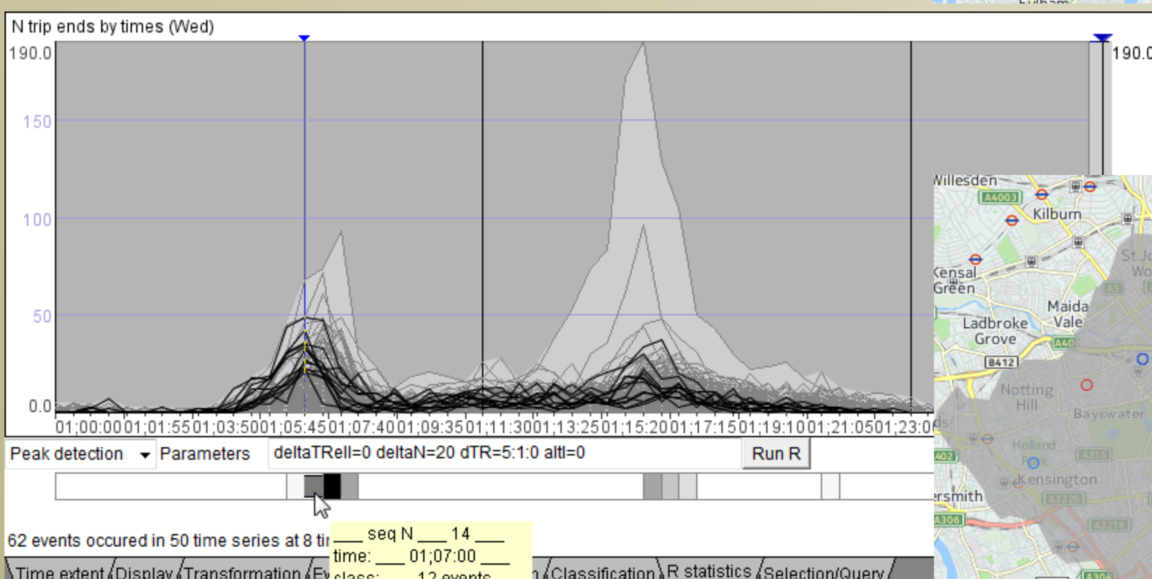
In a similar way, other time series features can be extracted and treated as spatial events.

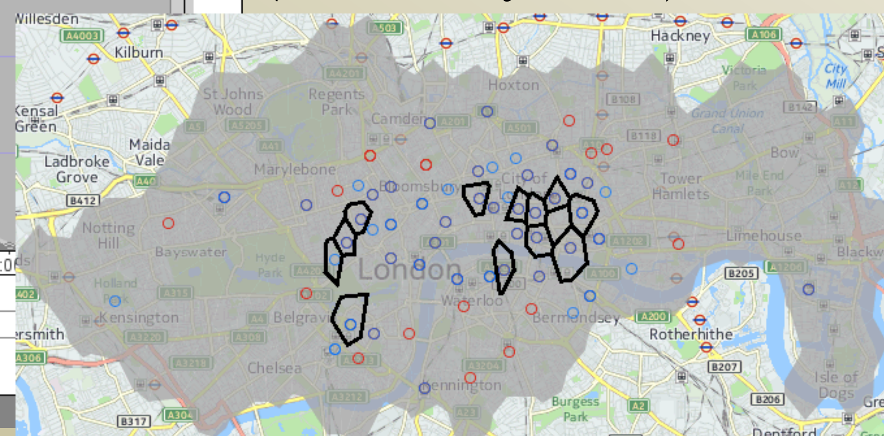
Peaks of trip starts (increase by >= 20)

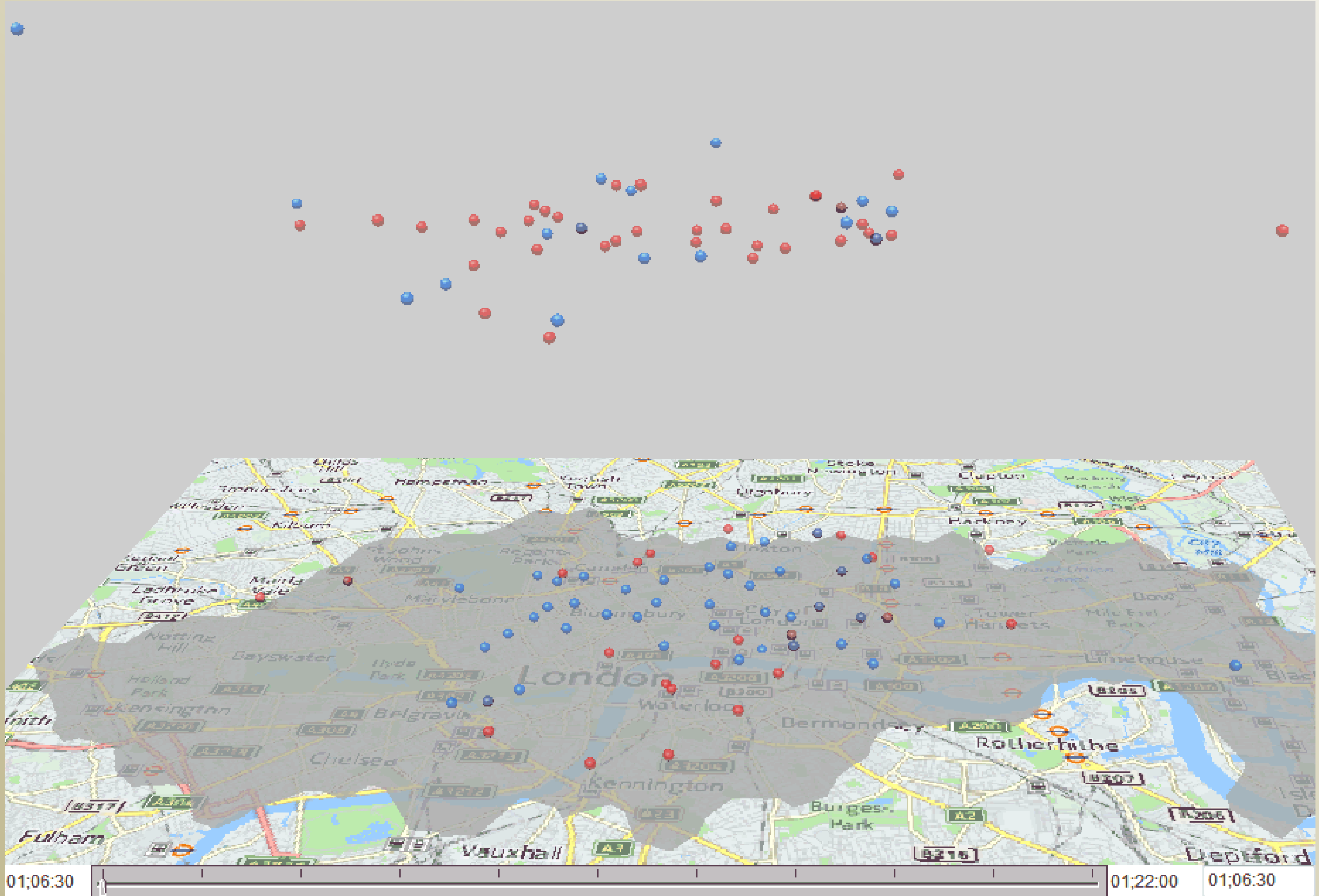


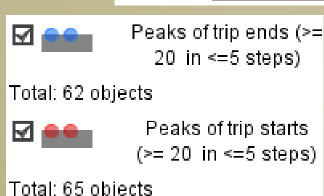


Peaks of trip ends (increase by >= 20)

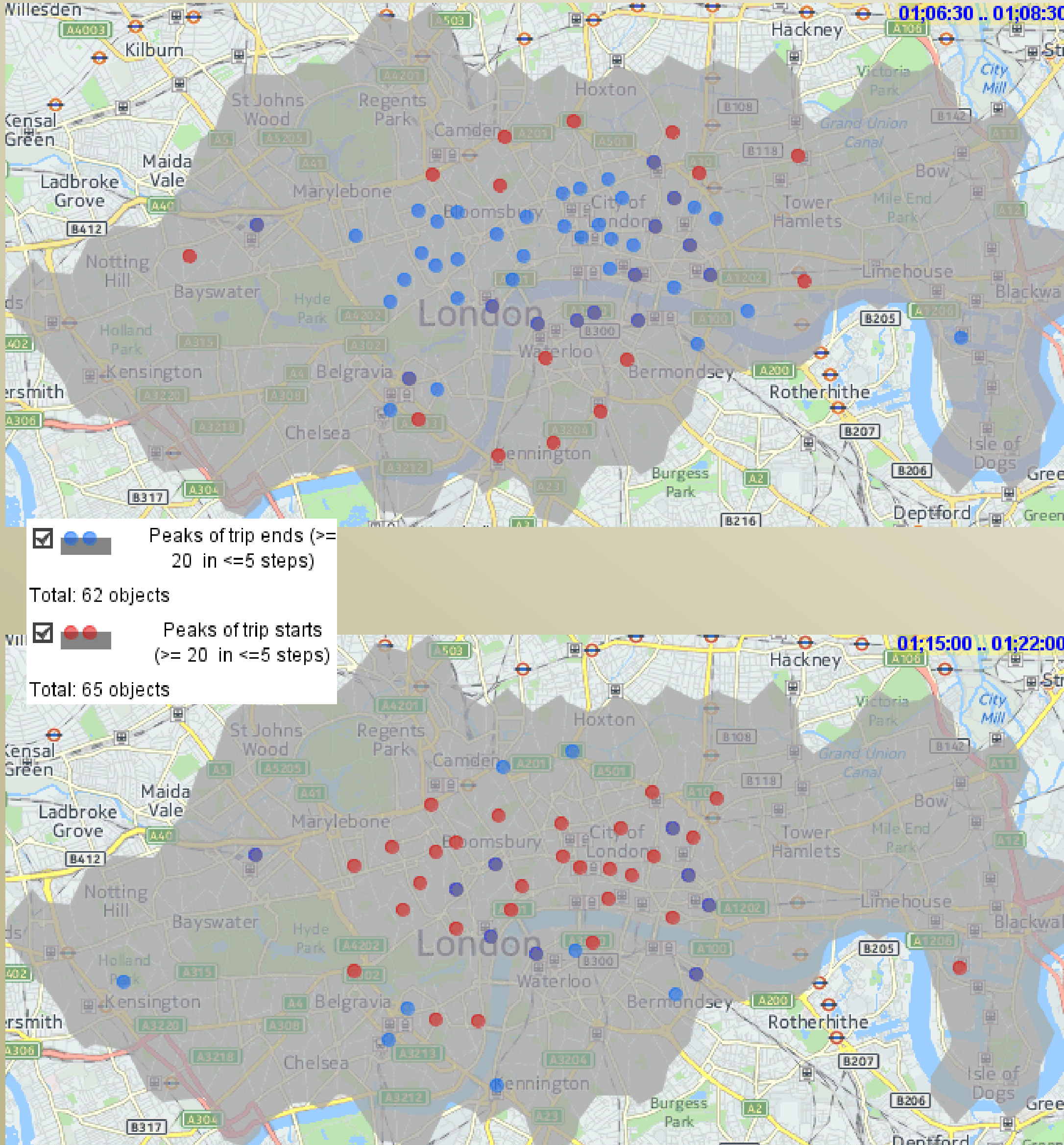








The STC shows the spatio-temporal distribution of the peaks of trip starts (red) and ends (blue). The violet colour of a ball means a spatio-temporal co-incidence of two different peaks.



There were 2 time periods of peaks in the numbers of the trip starts and ends: morning (6:30-8:30) and afternoon (15:00-18:00). These two screenshots of the map display show the spatial distribution of the morning (top) and afternoon (bottom) peak events. Many of the morning trip start peaks occur at railway stations, and the morning trip end peaks occur mostly in the centre. In the afternoon, an approximately inverse pattern is observed, while the number of afternoon peaks is smaller.

Where to read more:

http://dx.doi.org/10.1109/TVCG.2011.153

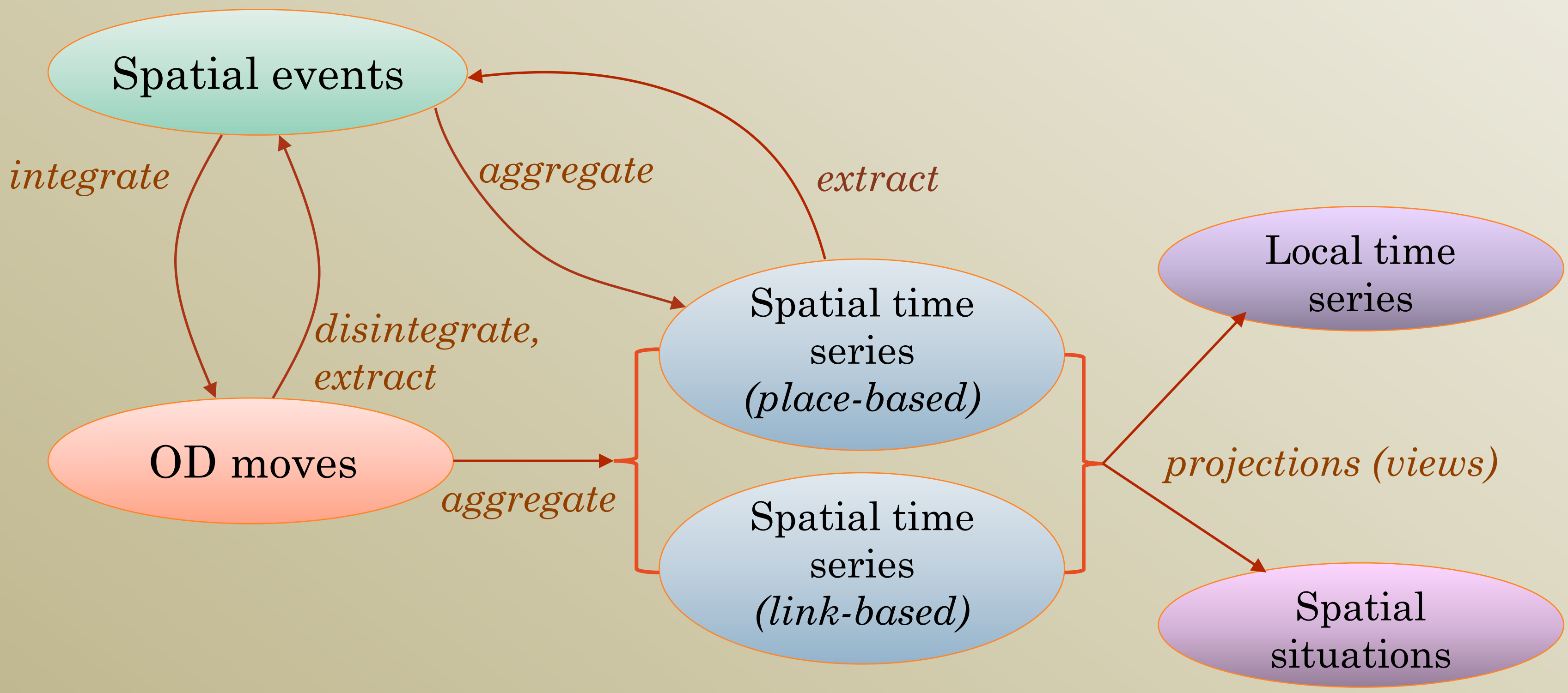
Gennady Andrienko, Natalia Andrienko,
Martin Mladenov, Michael Mock, Christian Pölitz

Identifying Place Histories from Activity Traces with an Eye to Parameter Impact

IEEE Transactions on Visualisation and Computer Graphics,
2012, v.18 (5), pp.675-688

Types of spatio-temporal data

and transformations between them



Purposes of data transformations

- **Aggregation**
  - Supports abstraction, gaining an overall view of characteristics and behaviour
  - Reduces large data
  - Simplifies complex data
- **Extraction** *of events, etc.*
  - Selects a portion of data relevant to a task, enables focusing
  - Allows dealing with complex data portion-wise
- **Integration, disintegration, projection**
  - Adapts data to analysis tasks

Analysis of spatial events and OD moves

Complementarity of DBC and spatio-temporal aggregation

- Task: analyse the spatio-temporal distribution of events/moves
  - Space and time are treated as containers of the events/moves
- Spatio-temporal aggregation:
  - Creates a data structure (spatial time series) that appropriately represents the distribution:
    - Space and time become referrers; attributes express containment of objects by spatio-temporal bins
  - However, it may conceal important features in the ST distribution
    - particularly, spatio-temporal concentrations of events and spatio-temporal coincidences of moves with close origins and/or destinations
    - Possible reasons:
      - Large bins: features become concealed by averaging
      - Small bins: features become dispersed over multiple bins

⇒ST aggregation need to be complemented by ST density-based clustering, which reveals concentrations and coincidences.

Questions?

Types of spatio-temporal data and transformations between them