EuroVis Workshop on Visual Analytics (2019)
C. Turkay and T. von Landesberger (Editors)

Contextualized Analysis of Movement Events

Siming Chen[1 †], Gennady Andrienko[1,2], Natalia Andrienko[1,2], Christos Doulkeridis[3] and Athanasios Koumparos[4]

[1]Fraunhofer IAIS, Germany
[2]City, University of London, UK
[3]University of Piraeus, Greece
[4]Vodafone Innovus S.A., Greece

Abstract

For understanding the circumstances, causes, and consequences of events that may happen during movement (e.g., harsh brake, sharp turn), it is necessary to analyze event context. The context includes dynamic attributes of the moving objects before and after the event and external context elements such as other moving objects, weather, terrain, etc. To explore events in context, we propose an analytical workflow including event contextualization, context pattern detection, and exploration of the spatio-temporal distribution of the detected patterns. The approach involves clustering of events based on the similarity of their contexts and interactive visual techniques for exploration of the distribution of the clusters in time, geographic space, and multidimensional attribute space. In close collaboration with domain experts, we apply our method to real-world vehicle trajectories with the purpose of identifying and investigating potentially dangerous driving behaviors.

1. Introduction

Modern movement tracking technologies allow position recording to be accompanied by collection of additional attributes. These attributes may describe properties of the moving objects (e.g. fuel level), attributes of its movement (e.g. speed), and parameters of the environment (e.g. temperature). Often recordings include specific events, such as harsh braking or cornering. Such events can be derived from raw positional data or detected by sensors.

Understanding the conditions in which movement events happen and/or their effects may be important in various applications. Thus, in businesses that rely on using vehicles, there is an interest in detecting bad driving habits as well as in identifying locations and times characterized by increased risk of crashes or other unwanted events. For such purposes, it is necessary to analyze the event *context*, which includes values of attributes characterizing the movement (e.g., speed, heading, etc.) and the environment (e.g., other moving objects, terrain, weather, etc.) before and after the events.

Previous researches dealing with movement events have been mostly focusing on extraction of various events from trajectories. Once extracted, the events were analyzed separately. We propose a new perspective for investigating movement events: *contextualized analysis*. The approach involves a procedure of *event contextualization*, which consists of (1) extracting segments of trajectories preceding and following the events, and (2) enriching these segments with values of additional attributes characterizing the movement and/or describing the external circumstances. The result consists of time series of attribute values, where the time references are relative with respect to the times of the events. Varying sampling rates in trajectories may require additional pre-processing operations such as interpolation. We apply visually driven cluster analysis to these time series to find and interpret repeating patterns. Then, the spatial and temporal distribution of these patterns is explored using interactive visual displays. The effectiveness of the proposed procedure has been tested with real data describing movement of trucks in Greece, addressing needs of industrial partners in a large project.

2. Related Work

Visual analytics of movement has been an area of intensive research in recent years [AAB*13]. The most common visualization techniques for movement data are maps, timelines, and space-time cube [Kra03]. To reflect the movement dynamics in maps, animation and small multiples are used [FPV*13]. The application areas of movement analytics are broad, including urban planning [FPV*13], social media research [CYW*16], ground transportation [AAC*17], air traffic analysis [AAGS19], sport analytics [AAB*17], studies of animal behavior [BJC*19], etc.

Extraction and analysis of movement events has been one of research directions in visual analytics of movement data. Andrienko et al. proposed a workflow including event extraction, density-based clustering for defining places of event concentration in space and time, and analyzing temporal patterns of event occurrence in these places [AAH*11b]. Wang et al. extract traffic jam events by exploring small multiples of jam propagation graphs [WLY*13]. These works did not involve analysis of event context apart from

† This research was partly funded by EU project Track&Know (780754)

DOI: 10.2312/eurova.20191124

event distribution in space and time. Several papers study multi-variate spatial-temporal data, including VIS-STAMP [GCML06] and Attribute Signature [TSH*14], but they do not specifically address movement events and trajectories in which the events occur.

In [AAH11a], a conceptual model has been proposed for representing a trajectory as a sequence of events, extracting events from trajectories and analyzing relationships between the events. Our current work is complementary, considering events in the context of trajectories and their attributes. Several papers dealt with dynamic attributes along trajectories. Scheepens et al. applied a multi-attribute filter and visualized the results in a density map [SWVdW*11]. Tominski et al. proposed a technique called Trajectory Wall, a 3D view in which similar trajectories are represented by bands stacked over a map background and coloured based on attribute values [TSAA12].

Our approach targets at bridging the gap between event detection, study of events distribution in space and time, and study of dynamic attributes describing event context.

3. Overview

Our data originate from tracking devices installed in commercial fleet vehicles. Our project partners provided a large data set covering one year of data across Europe. For designing and testing the approach, we have used a small subset consisting of 140,506 recorded positions of five trucks that moved in Greece over the period of 3 months. The data include vehicle description specifying the car type, size of its tank, etc., movement information consisting of time-stamped positions and values of direction, speed, engine status, odometer, fuel amount, and a few other attributes. In addition, some records contain codes of particular driving events, such as harsh brake, harsh acceleration, and harsh cornering. Partners are interested in using these data for characterizing driving behaviours, which includes understanding of the situational context in which the events occur. The context includes the character of the movement and the dynamics of the external conditions before and after the event occurrences.

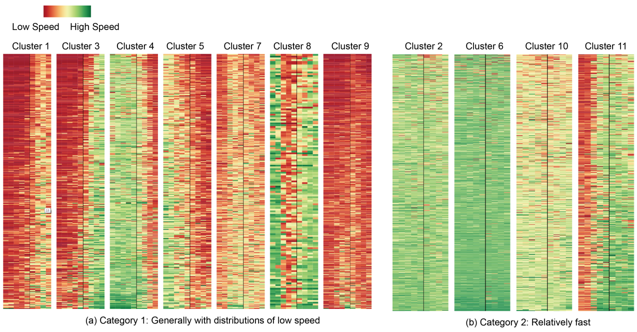
Obviously, it is neither feasible nor worthwhile to investigate the particular context of each individual event. What is interesting and valuable for the business partners is finding and interpreting repeated patterns of situations in which events occur. Once patterns are discovered, it needs to be checked if any of them are strongly associated with particular places and/or times. The existence of such associations may be taken into account in defining vehicle routes and/or temporal scheduling of their trips.

To achieve the goals, we propose a generic analytical workflow for contextualized analysis of movement events. The workflow is designed to work in a variety of applications where events in the context of dynamic position-related attributes are of interest. Before the analysis, the data undergoes pre-processing (such as cleaning) and enrichment with attributes describing external conditions, such as terrain, weather, type of road, type of territory, etc.

The analysis starts with selection of events of interest using a combination of query tools for selecting locations (e.g. only on the selected islands), times (e.g. only weekends in September), and event characteristics (e.g. harsh acceleration of a given magnitude). The result of this step is a list of events with their locations, times, and references to the moving objects from whose trajectories the events have been extracted. At the second step, the analyst selects the relevant attributes, the desired time window in relation to the event time, and the temporal resolution considering the sampling rate of the available data. In a case of varying sampling rates, interpolation or smoothing need to be applied. In the result of this step, each event is characterized by a vector of contextual attribute values. At the third step, repeated patterns are discovered by applying clustering to the vectors of all events with an appropriate similarity measure and the clustering technique. To support the clustering process and find a suitable number of clusters, we apply projection of cluster centers as proposed in [AAFW17] (see Fig. 2e). Attribute characteristics of the clusters are presented in visual displays for comparison and semantic interpretation. At the fourth step, the analyst investigates the spatial and temporal distributions of the clusters aiming at finding spatial and/or spatio-temporal "hot spots". At any step, the analyst may decide to return to one of the prior steps and change some of the choices made earlier.

4. Visual Analytics Workflow

We demonstrate the utility of the proposed workflow by investigating the "harsh braking" events in the context of speed attributes. The test dataset includes 5,058 harsh brake events. An event can be represented as $Event = (m, T, p, t)$, where m is an identifier of a moving object, T is an event type, p is a spatial position, and t is a time moment. Data pre-processing includes removal of unrealistic attribute values, such as truck speed values above 150km/h, taking into account the statistical distributions of the values.



(a) Category 1: Generally with distributions of low speed
(b) Category 2: Relatively fast

Figure 1: *Examples of clusters of contextualized events.*

4.1. Event Contextualization

For event contextualization, a suitable relative time window around the time of event occurrence needs to be chosen. Generally, the choice is domain-specific, depending on the character and velocity of the movements, character of the events, as well as temporal resolution of the data. Taking into account the sampling rate of our data, we tested the windows ± 3, ± 5, and ± 10 minutes. After a discussion with the domain experts, we chose ± 5 minutes. The window length is divided into equal time steps. For different analysis purposes, the granularity and the time window size may be different.

The contextualization consists of supplying each event with values of selected relevant attributes for each time step within the relative time window. The output is the $ContextualizedEvent =$

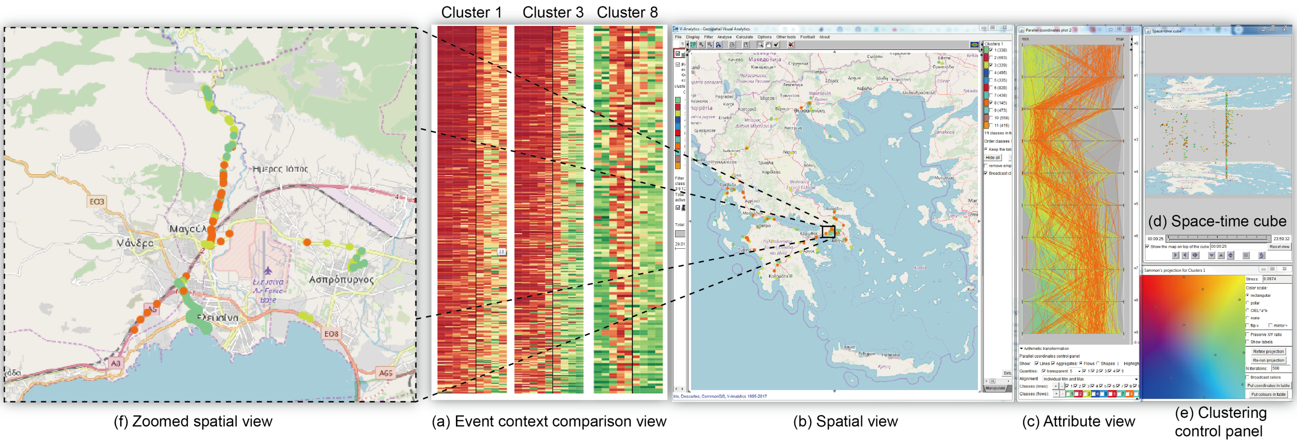
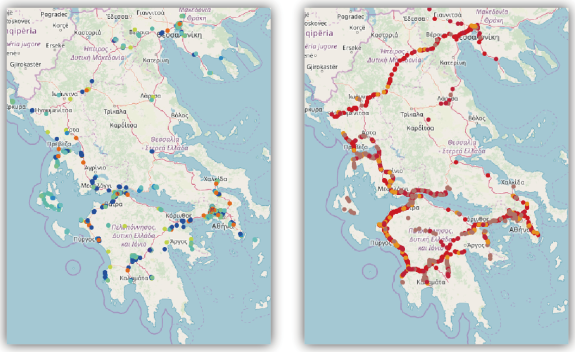


Figure 2: *Visual analytics interface, including (a) Event Context Comparison View, (b) and (f) Spatial View, (c) Attribute Parallel Coordinates View, (d) Space Time Cube, (e) K-Means Clustering Control Panel.*



(a) Category 1: Intra-city (Passing City) Events (Cluster 1,3,4,5,7,8,9)

(b) Category 2: Inter-city Events (Cluster 2,6,10,11)

Figure 3: *Comparison of spatial distributions of events with distinct context patterns.*

$(m, T, p, t, Attrs[-\Delta t_1, +\Delta t_2])$, where $[-\Delta t_1, +\Delta t_2]$ is a relative time interval with respect to the event time (which is treated as zero) and $Attrs[-\Delta t_1, +\Delta t_2]$ is the corresponding time series of values of one or more attributes. To obtain attribute values for the time steps from the original data, interpolation and/or aggregation operators are used. Thus, if the time intervals between the data records are smaller than the chosen time step, the means or medians of the original values can be taken. In our example, both Δt_1 and Δt_2 are equal to 5 minutes and the time step is 1 minute length; hence, each event is characterized by a vector consisting of 10 speed values.

4.2. Context Pattern Discovery

After contextualizing the events, clustering by similarity is applied to the event contexts, i.e., vectors of attribute values. Taking into account the dimensionality of the vectors and the statistical distribution of the values, we chose kMeans as the clustering method and Euclidean distance as the similarity measure. We iteratively performed clustering varying the cluster number from 5 to 20. To choose a suitable number of clusters, we looked after each run at the statistics of the distances of the cluster members to the cluster centers and at a 2D projection of the cluster centers obtained by applying the Sammon's projection method to the vectors of the mean attribute values for the clusters (Figure 2e). The former reflects the intra-cluster variation (a high variation suggests that the number of clusters should be increased) and the latter shows how different the clusters are (small distances between some points suggest that the respective clusters do not substantially differ, and thus the number of clusters is excessive). Following this procedure, we chose k=11.

By applying a continuous color scale to the projection space, we assign colors to clusters reflecting their similarity. These colors can be used for representing the clusters in other displays (map, parallel coordinates, space-time cube etc.)

Each cluster can be seen as a realization (i.e., a set of instances) of some context pattern. To enable interpretation and comparison of the cluster-specific patterns, we visualize each cluster as a matrix where the rows correspond to the cluster members (i.e., the events) and the columns correspond to the attribute values at the time steps from $-\Delta t_1$ to $+\Delta t_2$) (Figure 1). The values are represented by colors. In our example, red is used for low speeds and green for high. The rows are ordered from top to bottom according to their average speeds. The matrices of different clusters are juxtaposed for comparison. They have the same overall height; hence, their row heights differ due to the differences in the cluster sizes. In our test study, two groups of patterns were found: low-speed (Figure 1a) and high-speed (Figure 1b). Within these groups, one can identify patterns of speeding up (cluster 1, 3, 11), slowing down (cluster 4, 5), and distinct variation patterns (cluster 7, 8, 10). The domain experts found these patterns understandable and meaningful.

4.3. Exploration of the Spatial and Temporal Distribution

Following the overview of the clusters of contextualized events (Figure 2a), the next step in the analysis is to explore the spatial (Figure 2b), temporal (Figure 2d) and attribute patterns of the events (Figure 2c). Thus, Figure 3 shows the spatial distributions of the two large groups of context patterns visible in Figure 1. These groups of patterns correspond to intra-city (Figure 3a) and inter-city movements (Figure 3b). This makes sense as the speeds on the inter-city roads are generally higher. For the exploration of the spatio-temporal distribution, we use the space-time cube in which the absolute time range is transformed to the positions within the 24-hours daily cycle (Figure 5). The two groups have different tem-

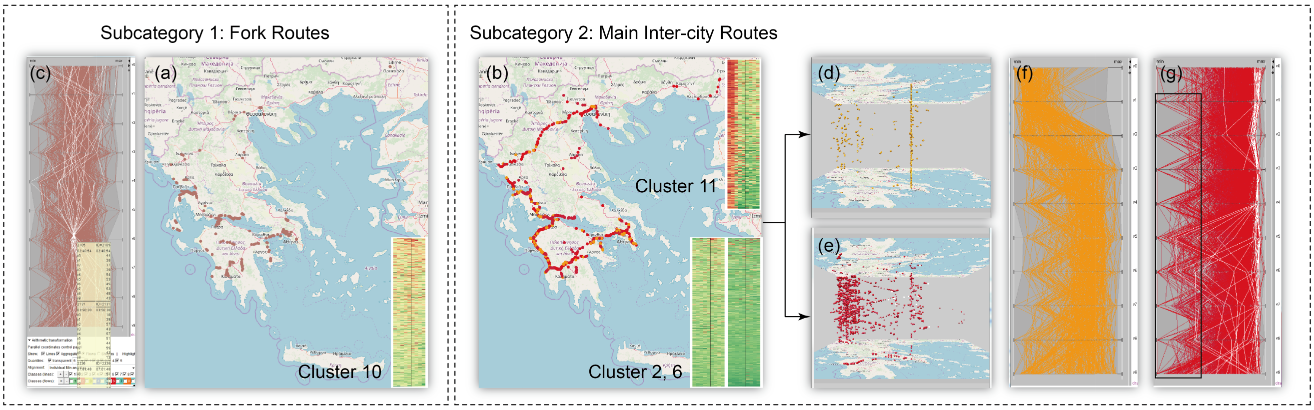
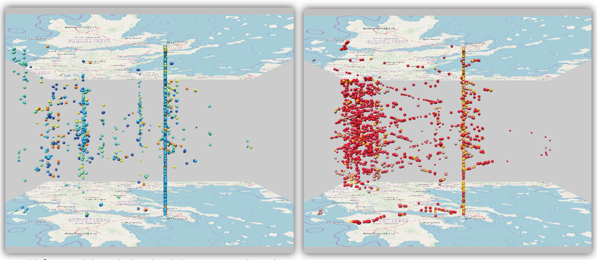


Figure 4: *Exploration of the clusters of contextualized events from the inter-city group (Fig. 3b). One of the clusters is spatially associated with fork layouts of roads (a) and the others with main roads (b). The latter form two sub-groups with patterns of speed increase (d,f) and speed dropping (e,g). Sporadic speed drops on a high-speed road may be dangerous.*

poral patterns. The intra-city movement events happen more often at the day time in most locations (Figure 5a) while the inter-city events last till late in the night (Figure 5b).



(a) Category 1: Intra-city (passing city) events space-time cube (Cluster 1,3,4,5,7,8,9)

(b) Category 2: Inter-city events spacee-time cube (Cluster 2,6,10,11)

Figure 5: *Comparison of the spatio-temporal distributions of two event groups. The event times have been transformed to the relative dimes within the 24-hours daily cycle, which is represented by the vertical dimension of the cube.*

The intra-city group (Figure 3a) includes patterns of increasing speed, decreasing speed, and variable speed before and after the harsh brake events. For investigating the increasing speed patterns (Figure 2a), we use a parallel coordinates plot where the vectors of attribute values are represented by lines colored according to the cluster membership (Figure 2c). Cluster 1 and 3 represent harsh brake events that happened in the course of slow movement before the speeds started to increase (Figure 2a, c). These patterns may correspond to final parts of movement in traffic jams. Cluster 8 is different. Here, speeds decreased from normal to slow, and then harsh breaking happened just before accelerating to normal speeds. The domain experts interpreted these patterns as probably corresponding to stopping at red lights on street intersections or encountering slow moving vehicles ahead. Map zooming (Figure 2b) highlights regions where these patterns occurred (Figure 2f). Based on this information, it may be recommended to avoid the identified regions at specific times or to drive there with special attention.

In the inter-city group of clusters (Figure 1b), the domain expert found two subcategories (Figure 4a, b). The events of cluster 10 happened mostly at fork layouts of roads. The context pattern shows high variation within the speed value range (Figure 4c). The events of the other clusters occurred mostly on main roads. This group of clusters can be further divided into two sub-groups with different patterns of speed values and temporal distribution (Figure 4d,e). Cluster 11 consists of events that happened during speeding up (Figure 4f), while clusters 2 and 6 include events that occurred in the context of generally fast movement with sudden speed drops and even stops (Figure 4g). The domain expert noted that such harsh breaking events may be especially dangerous on high-speed inter-city roads.

5. Discussion and Conclusion

Analysis of circumstances in which certain movement events happen may be useful for various reasons, e.g., when the events are unwanted and should be avoided, or the other way around. We propose a data transformation and a visual analytics workflow for this kind of analysis. The practical example of data analysis presented in this paper has been limited to a small dataset. Its purpose was to test the main idea of the approach from the perspective of the technical feasibility and utility for the domain experts. In the further work, we shall scale up the approach to large data volumes with additional environmental information and investigate the possibilities for extending it to real time analysis of streaming data.

Specifically, this includes development of a database back-end for out-of-core event extraction and contextualization. We are also designing novel visualization techniques for displaying contextual descriptors of large amounts of individual events in an aggregated form and for giving a proper overview of clusters of contextualized events. Another direction of our work is analysis of multivariate contexts of events. The computational components involved in our approach are applicable to multivariate data while the visualization part needs to be developed further. We expect that sorting according to different criteria [MGKH07] can be helpful. To deal with large amounts of data that may be too heavy for direct application of clustering, we plan to use an approach in which clustering is applied to manageable data samples followed by visually driven generation of classifiers for assigning new members to the clusters defined [AAR*09].

References

[AAB*13] Andrienko G., Andrienko N., Bak P., Keim D., Wrobel S.: *Visual analytics of movement*. Springer Berlin Heidelberg, 2013. 1

[AAB*17] Andrienko G., Andrienko N., Budziak G., Dykes J., Fuchs G., von Landesberger T., Weber H.: Visual analysis of pressure in football. *Data Mining and Knowledge Discovery 31*, 6 (Nov 2017), 1793–1839. doi:10.1007/s10618-017-0513-2. 1

[AAC*17] Andrienko G., Andrienko N., Chen W., Maciejewski R., Zhao Y.: Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems 18*, 8 (Aug 2017), 2232–2249. doi:10.1109/TITS.2017.2683539. 1

[AAFW17] Andrienko G., Andrienko N., Fuchs G., Wood J.: Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE Transactions on Visualization and Computer Graphics 23*, 9 (Sep. 2017), 2120–2136. doi:10.1109/TVCG.2016.2616404. 2

[AAGS19] Andrienko N., Andrienko G., Garcia J. M. C., Scarlatti D.: Analysis of flight variability: a systematic approach. *IEEE transactions on visualization and computer graphics 25*, 1 (2019), 54–64. doi:10.1109/TVCG.2018.2864811. 1

[AAH11a] Andrienko G., Andrienko N., Heurich M.: An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science 25*, 9 (2011), 1347–1370. doi:10.1080/13658816.2011.556120. 2

[AAH*11b] Andrienko G., Andrienko N., Hurter C., Rinzivillo S., Wrobel S.: From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *Proc. IEEE VAST* (2011), pp. 161 – 170. doi:10.1109/VAST.2011.6102454. 1

[AAR*09] Andrienko G., Andrienko N., Rinzivillo S., Nanni M., Pedreschi D., Giannotti F.: Interactive visual clustering of large collections of trajectories. In *2009 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2009), pp. 3–10. doi:10.1109/VAST.2009.5332584. 4

[BJC*19] Buchmüller J., Jäckle D., Cakmak E., Brandes U., Keim D. A.: MotionRugs: Visualizing collective trends in space and time. *IEEE transactions on visualization and computer graphics 25*, 1 (2019), 76–86. 1

[CYW*16] Chen S., Yuan X., Wang Z., Guo C., Liang J., Wang Z., Zhang X. L., Zhang J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan 2016), 270–279. doi:10.1109/TVCG.2015.2467619. 1

[FPV*13] Ferreira N., Poco J., Vo H. T., Freire J., Silva C. T.: Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec 2013), 2149–2158. doi:10.1109/TVCG.2013.226. 1

[GCML06] Guo D., Chen J., MacEachren A., Liao K.: A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Trans. Vis. Comput. Graph. 12*, 6 (2006), 1461 –1474. doi:10.1109/TVCG.2006.84. 2

[Kra03] Kraak M.: The space-time cube revisited from a geovisualization perspective. In *The 21st international cartographic conference* (2003), International Cartographic Association (ICA). 1

[MGKH07] Matkovic K., Gracanin D., Konyha Z., Hauser H.: Color Lines View: An approach to visualization of families of function graphs. In *2007 11th International Conference Information Visualization (IV '07)* (July 2007), pp. 59–64. doi:10.1109/IV.2007.35. 4

[SWVdW*11] Scheepens R., Willems N., van de Wetering H., Andrienko G., Andrienko N., van Wijk J. J.: Composite density maps for multivariate trajectories. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2518–2527. doi:10.1109/TVCG.2011.181. 2

[TSAA12] Tominski C., Schumann H., Andrienko G., Andrienko N.: Stacking-based visualization of trajectory attribute data. *IEEE Trans. Vis. Comput. Graph. 18*, 12 (2012), 2565–2574. doi:10.1109/TVCG.2012.265. 2

[TSH*14] Turkay C., Slingsby A., Hauser H., Wood J., Dykes J.: Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 2033–2042. doi:10.1109/TVCG.2014.2346265. 2

[WLY*13] Wang Z., Lu M., Yuan X., Zhang J., van de Wetering H.: Visual traffic jam analysis based on trajectory data. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2159–2168. doi:10.1109/TVCG.2013.228. 1