Visual Analytics in Urban Computing: An Overview

Yixian Zheng, Wenchao Wu, Yuanzhe Chen, Huamin Qu, *Member, IEEE*, and Lionel M. Ni, *Fellow, IEEE*

Abstract—Nowadays, various data collected in urban context provide unprecedented opportunities for building a smarter city through urban computing. However, due to heterogeneity, high complexity and large volumes of these urban data, analyzing them is not an easy task, which often requires integrating human perception in analytical process, triggering a broad use of visualization. In this survey, we first summarize frequently used data types in urban visual analytics, and then elaborate on existing visualization techniques for time, locations and other properties of urban data. Furthermore, we discuss how visualization can be combined with automated analytical approaches. Existing work on urban visual analytics is categorized into two classes based on different outputs of such combinations: 1) For *data exploration and pattern interpretation*, we describe representative visual analytics tools designed for better insights of different types of urban data. 2) For *visual learning*, we discuss how visualization can help in three major steps of automated analytical approaches (i.e., cohort construction; feature selection & model construction; result evaluation & tuning) for a more effective machine learning or data mining process, leading to sort of artificial intelligence, such as a classifier, a predictor or a regression model. Finally, we outlook the future of urban visual analytics, and conclude the survey with potential research directions.

Index Terms—Urban computing, visual analytics, visualization, visual learning, spatio-temporal, multivariate

1 Introduction

With the development of science and technology, urbanization process has been accelerating worldwide, which on one hand improves people's life quality, on the other hand gives rise to serious problems, such as environmental pollution, traffic congestion and ever-increasing energy consumption. As data collection becomes easier and cheaper, a wider variety of big data in urban space, such as human mobility data and air quality data, are generated and become available. These data make it possible to tackle challenges that we are facing and help build smarter cities. For instance, we can analyze urban traffic congestions based on GPS trajectories collected from taxis [1] and explore causes of air pollution by correlating air quality data with other related data sources, such as road network, traffic, and point of interests (POIs) [2]. The findings could be used to support decision making and help better formulate city planning for the future. Inspired by the vision of better cities, urban computing has drawn more and more attentions of researchers from different fields, who aim to unlock the power of knowledge from big and heterogeneous data collected in urban context and apply this powerful information to tackle problems challenging us at present [3].

Although the term "urban computing" was coined and first used in 2003 by Eric Paulos [4] and many researchers [5], [6], [7], [8] have been working on it over years, there are still quite a few issues which have not been addressed satisfactorily. Recently, Zheng et al. [3] presented a survey on urban computing, which introduced general framework, key research problems, methodologies, and applications mainly based on automated data mining approaches. However, as known to all, urban computing is a multi-disciplinary research field, where computer science meets conventional city-related areas, such as civil engineering, transportation, economics, energy engineering and environmental science, which usually leads to complex analytical tasks for real world applications. Therefore, a fully automatic analysis is difficult, often requiring considerable experience and profound knowledge in various fields. It is important to include human perception in the data exploration process and combine the flexibility, creativity and domain knowledge of human beings with enormous storage capacity and computational power of today's computers.

Visualization, the study of transforming data and information into interactive visual representations [9], provides an effective way to integrate humans in a data exploration process, applying their perceptual abilities to the target datasets and leveraging their domain knowledge to guide the exploration. Furthermore, visual analytics combines automated analysis with interactive visualization for effective understanding, reasoning and decision making on the basis of a very large and complex dataset [10]. Though the importance of visual analytics has been recognized, especially for urban computing [3], it remains a vague concept with many questions still pending. For example, what role can visualization play in urban computing? What are the representative visual designs in this domain? How can we combine visualization with automated approaches of data mining or machine learning, and how does an urban visual analytics system work?

In order to address these issues, in this article, we focus on visual analytics in urban computing (hereafter referred

• *Y. Zheng, W. Wu, Y. Chen, and H. Qu are with the Hong Kong University of Science and Technology, Kowloon, Hong Kong. E-mail: {yixian.zheng, wenchao.wu, ychench}@connect.ust.hk, huamin@cse.ust.hk.*
• *L.M. Ni is with the University of Macau, Taipa, Macau, China. E-mail: ni@umac.mo.*

Manuscript received 28 Dec. 2015; revised 28 Apr. 2016; accepted 15 May 2016. Date of publication 6 July 2016; date of current version 28 Oct. 2016. Recommended for acceptance by D. Zhang.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2016.2586447

to as *urban visual analytics*) and discuss several important issues from the perspective of visualization. We hope this article will help the community better understand and explore this area, therefore guide future work that can eventually lead to better cities. The paper hereunder is organized as follows: The data types frequently used in urban visual analytics are first discussed in Section 2. Then Section 3 elaborates on the various visualization techniques for spatial, temporal and other properties of urban data. In Section 4, based on the framework of urban visual analytics, we discuss how visualization can be combined with automated approaches to enhance understanding and mining of urban data. Finally, this article is concluded and future work is highlighted in Section 5.

2 Urban Data

In real-world applications, solving urban challenges usually needs to consider a broad range of factors, which requires a clear picture of what data can be leveraged in urban context. Meanwhile, different types of urban data demand different visualization and analysis methods. In this section, we categorize the frequently used urban data types in the field of visualization into six categories (i.e., human mobility data, social network data, geographical data, environmental data, health care data and others), and discuss the common properties of urban data which need to be taken into account for visual analysis. Table 1 summarizes the respective attributes of different data types and representative existing datasets.

2.1 Frequently Used Data Types

2.1.1 Human Mobility Data

In recent years, human mobility data is one of the most frequently used data types in urban visual analytics. It can facilitate the study of social and community dynamics, which is an important issue for many practical applications in the modern society. Based on different data sources, human mobility data can be further categorized into traffic data, commuting data, mobile phone data and geo-tagged social media data.

- **Traffic data:** refers to the type of data generated and collected by sensors in traffic vehicles (e.g., taxis, buses, metros, trains, vessels and planes) or monitors installed along the roads (e.g., loop sensors, surveillance cameras) [11].
  1) *Vehicle-based traffic data* records positions of vehicles from time to time and form a series of trajectories with temporal (i.e., timestamp) and spatial (i.e., longitude and latitude) information. Other information accompanying with trajectories, such as instant speed and heading directions, can also be collected. Compared with other traffic data, vehicle-based data can provide more details of movement, while the coverage of data still highly depends on the distribution of the probing vehicles and it is challenging to recover the citywide social and community dynamics based on limited data. In recent years, efforts have been made to deal with the limitations of data and support various applications [5].
  2) *Loop sensors* are usually embedded in pairs on major roads to detect the time interval that a vehicle travels across two consecutive sensors. Based on the data collected by loop sensors, we can easily calculate the travel speed as well as traffic volume on roads to perform a network analysis of traffic [12]. However, one obvious limitation of loop sensors data is the limited coverage, nor can it tell us any details about how a vehicle travels on a road.
  3) Surveillance cameras are widely deployed in urban areas nowadays, generating a huge volume of images and videos. This type of traffic data, called *surveillance data*, provides a visual ground truth of traffic conditions. However, it is still a challenging task to automatically extract information, such as traffic volume and flowrate, from these images and videos. Thus, currently surveillance data only provides a way to monitor citywide traffic conditions manually, which is obviously inefficient.
- **Commuting data:** is a type of data recording people's regular movement in cities. Among various data studied in the visualization community, card-swiping data is a typical example of this type. Nowadays, in a modern city, passengers often use their personalized RFID cards to tap on card readers on buses or metro station entries to enter/exit the public transportation system, thus generating a huge amount of records of passenger trips in a public transportation system. Each trip record includes an anonymous card ID, tap-in/out stops, time, fares for this trip and transportation type (i.e., bus or metro). This type of data can be used not only to improve the public transportation in a city [13] but also analyze citywide human mobility patterns [14].
- **Mobile phone data:** refers to data records of all exchanges (e.g., phone calls, messages, internet) between mobile phones and cell stations collected by telecom operators. In addition to communication information, this type of data provides locations of users from time to time based on cell stations, which offers unprecedented information resources to study human mobility [15], [16], [17] in terms of the large coverage and fine-grained resolution of urban population.
- **Geo-tagged social network data:** refers to a part of posts (e.g., blogs, tweets) through social networks which are tagged with geo-information. The availability of spatial and temporal information in social media can help us better understand people's activities [15], [18], [19]. Although rich information contained in such type of data makes it popular and interesting for human mobility analysis, the major challenges lie in the sparsity and uncertainty of data.

2.1.2 Social Network Data

Nowadays, social network becomes one of the most popular means of communication, generating huge amount of data, called social network data. Besides geo-information discussed previously, this type of data contains valuable

TABLE 1
Frequently Used Urban Data

Data Category			Data Property		Type N	Type C	Type T	Representative Datasets
Human mobility data	Traffic data	Vehicle-based	Time		√			Taxi [1] [23] [24] [25] [26] [27] [28] [29] [30] [31], bike [3] [32], aircraft [33] [34], train [35], vessel [34] [36] [37]
			Location		√			
			Velocity		√			
			Direction		√			
			Vehicle type			√		
			Vehicle ID		√			
		Loop sensors	Cell information	ID			√	Loop sensors data in Beijing [12]
				Location	√			
				Direction	√			
			Vehicle record	Time	√			
				Plate number			√	
				Plate color		√		
				Cell/Lane ID	√			
				Speed	√			
		Surveillance	Time		√			Tunnel surveillance video [38], traffic video clips [39]
			Location		√			
			Image/Video					
	Commuting data	RFID card data	Card ID				√	Smart card data in Singpore [13] [40]
			Tap-in/out stops		√			
			Tap-in/out time		√			
			Transportation type			√		
	Mobile phone data		User ID				√	Telco data in Guangzhou [17], telecommunication data in Abidjan [16] and Rome [41]
			Time		√			
			Base station ID		√			
			Base station location		√			
	Geo-tagged social network data		(refer to social network data)					Weibo [18] [42], twitter [19], flickr [15]
Social network data			Time		√			Twitter [43] [44] [45] [46] [47] [48] [49] [50] [51] [52], youtube comments [53]
			Content				√	
			User profile					
			Post type			√		
			Geo-tag		√			
Geographical data	Road network		Intersections	ID			√	Manhattan [54] [55], Boston [56], Beijing [39]
				Location	√			
			Road segments	Start intersection			√	
				End intersection			√	
				Condition	√	√		
	Transporation network		Transit routes	Start stop			√	Bus [16], subway [57] [58], public transportation system [13]
				End stop			√	
			Stop facilities				√	
			Schedule information		√			
	POI		Location		√			Foursquare [59] [60]
			Address				√	
			Name				√	
			Category			√		
Environmental data	Environment monitoring		Time		√			Meteorological data [61], river water quality [62], air pollution [2] [63], satellite remote sensing data [64]
			Location		√			
			Indices		√			
	Energy consumption		Time		√			UK domestic energy consumption [65] [66], gas consumption [67]
			Appliance			√		
			Consumption		√			
			User ID				√	

Note that open data sources are highlighted with underlines, and N,C,T are short for numerical, categorical and textual respectively.

information in two aspects. On one hand, analyzing communications between users enables us to study the relationship among different people as well as the social structure of a certain community [20]. On the other hand, the user-generated social media, such as texts, photos and videos, contain rich information about a user's interests and characteristics, which provides references for researches on various social issues, like evolving of public attention on topics of social media [21], and spreading of anomalous information [22].

2.1.3 Geographical Data

Geographical data is a fundamental data type in urban visual analytics which provides basic structure as well as semantic information for urban computing scenarios. In the field of visualization, *road network data*, *transportation network data* and *POI data* (point of interest) are frequently used data of this type.

- **Road network data:** is usually in the structure of a graph that is comprised of a set of edges and nodes, representing road segments and intersections respectively. Each node is described by a unique set of geographical coordinates, while each edge is associated with other related properties, such as length, speed limit, type of road and number of lanes.
- **Transportation network data:** includes transit routes and stop facilities of the bus and metro network which is modeled as a directed graph. Each stop facility is described with ID, geographical coordinates and related edge connection in the network. In addition, schedule information is often included with a timetable showing when each bus/metro leaves its starting terminal, and reaches each stop along its transit route.

- **POI data:** depicts related information of facilities, such as restaurants, shopping malls, parks, airports, schools and hospitals in the city. Each facility is usually described by a name, address, category and a set of geographical coordinates.

2.1.4 Environmental Data

In recent years, accelerating urbanization has led to serious environmental problems worldwide, such as severer environmental pollution and ever-increasing energy consumption. Many studies on urban visual analytics have been dedicated to analyze related datasets to tackle the environmental problems. These datasets can be categorized into two classes, environment monitoring data and energy consumption data. The former includes meteorological data (e.g., temperature, humidity, sunshine duration and weather conditions) [69], [70], air pollution data [2], [61], water quality data [62] and satellite remote sensing data [64]; the latter records consumptions of electricity [65], gas [67] etc., which can help to evaluate and optimize energy usage in a city by detecting correlations [66] and predicting peak loads of demand [67].

2.1.5 Other Related Data

As we all know, modern cities are integrated and comprehensive units, thus there are quite a few other related urban data in addition to those data types mentioned above, such as health care data [71], [72], [73], [74], [75], public utility service data [76], economy data [77], [78], [79], [80], education data [81], [82], manufacturing data [83] and sports data [84], [85], [86], [87], [88].

With the development of computing and data technology, more and more urban data will become available, which enables us to build a better and smarter city in the near future. Meanwhile, the increasing complexity and heterogeneity of urban data will bring great challenges and call for more advanced data analysis technologies. Urban visual analytics, discussed in the following parts, is definitely a step forward towards this objective.

2.2 Properties

Based on existing works in urban visual analytics, three fundamental properties, time (when), space (where) and object (what) can be extracted from urban data. They are elementary components for telling a full story under urban context and can help structure the information domain. The three basic properties are briefly depicted as follows.

Time, mathematically speaking, is a continuous or discrete linearly ordered set consisting of time instants or time intervals, jointly called *time units* [89]. Meanwhile, the time is not only a linear sequence but also includes inherent cycles, like iterations of seasons, weeks and days. This property provides temporal information which is essential to organize urban data for an efficient and effective analysis.

Space can be regarded as a set of locations. This is another common property to support the analysis of urban data. The existing ways of specifying locations in space can be summarized as follows [89]: a) *coordinate-based referencing:* refers to tuples of numbers representing the distance to certain reference points or axes; b) *division-based referencing:* refers to compartments of an geometric or semantic-based division of space; c) *linear referencing:* refers to relative positions along linear reference elements, such as streets, rivers and trajectories.

Objects include a set of physical and abstract entities. Objects, which are often analyzed in urban visual analytics, can be classified into three types according to their spatial and temporal properties. A spatial object refers to an object having a certain position in space, such as vehicles, persons and facilities, while a temporal object refers to an object existing in a certain period of time, which can also be called event. And a spatio-temporal object represents an object with a specific position in both space and time domains. In addition, other related attributes are often associated with objects for a comprehensive analysis.

3 Visualization of Urban Data

In recent years, a few open tools have become available for data visualization, such as Google Charts [90], Datawrapper [91], Baidu Echarts [92], Many Eyes [93] and Tableau [94]. However, most of them are too general to meet practical needs and support complex analytical tasks of real-world applications. Therefore, in this section, we survey the existing visualization techniques for urban data. Specifically, we discuss how and what visual channels can be used to visualize time, locations and other properties, which can facilitate better understanding and guide future work.

3.1 Visualization of Time

Time is one of the most important properties of urban data. The question is: How can time be presented visually? In this section, we focus on the "*time*" that is regarded as the measure by which urban data can be ordered from the past through the present into the future. We consider the measure of time durations and intervals between data items as numerical properties whose visualization will be discussed in Section 3.3.

There are various ways of mapping time to visual variables [68] as shown in Fig. 1. Under urban context, the *axis-based design* is the most popular method thanks to its simplicity and understandability. To visualize linear time, the classic method is charts, such as line chart (Fig. 3a) and stacked graph (Fig. 3b), where time is mapped to a horizontal axis and time-dependent attributes mapped to a vertical axis. Thus, the peaks or valleys of variable evolutions over time can be indicated clearly. Moreover, in order to emphasize the cyclic character of time, a clock-like circular time axis is often adopted. Fig. 3c shows an example of circular time axis which visualizes hotness of 96 human activities during 24 hours of a day (one activity per ring) [95]. The radial layout can help to reveal potential periodic patterns more intuitively.

In addition, the axis-based design presents absolute time precisely, while color and connection are two common methods to interpret time of urban data relatively [64], [96], [97]. Examples of presenting chronological order of data using color and connection are given in Fig. 4. Note that color [64] and connection [98], [99] can also be used separately. Although these two methods are widely used, their common limitation is the scalability which is caused by the limited

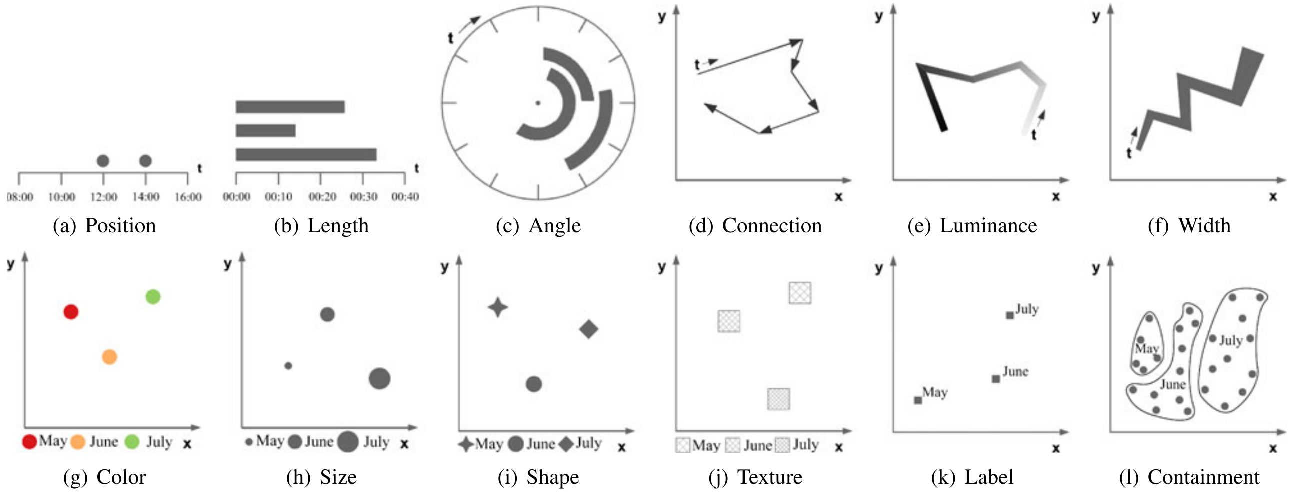


Fig. 1. Examples of mapping time to visual variables [68].

capability of human's eyes in distinguishing different colors and the potential clutter of connections respectively. Therefore, they are best used in combination with details-on-demand interactions.

Besides the static visualizations discussed above, temporal information can also be conveyed through a dynamic visual representation, which results in visualizations that change over time automatically (e.g., animation) [100]. However, as demonstrated by Robertson et al. [101], animation is generally not effective for analytical tasks due to the limitation of human short-term memory.

3.2 Visualization of Locations

With increasing availability of location-acquisition technologies, lots of urban data are collected with geographical locations. As discussed in Section 2.2, locations in urban context are often specified by three means: coordinate-based referencing, division-based referencing and linear referencing. Here we discuss the corresponding visualization techniques for these three types of locations, namely point-based visualization, region-based visualization and line-based visualization.

3.2.1 Point-Based Visualization

Point-based visualization is the most direct and intuitive type of visualization to present and analyze locations, as most locations are commonly recorded based on geographical coordinates in raw data. The basic idea of this type of techniques is to place points individually within spatial context (e.g., on a map). Each point represents an object [35], [102] or event [15], [24], [103], and visual channels (e.g., color, size) of these points encode related information (e.g., status of objects, category of events).

For instance, in TaxiVis [24], points in different colors are used to represent pickups and drop-offs of taxi trips in Manhattan to identify regular patterns as well as anomalies. As shown in Fig. 2, we can clearly see that, during 8-10 am, there were almost no taxis trips along 6th Avenue which turns out to be the holding place of Five Boro Bike Tour where traffic was blocked.

In addition to marking a spot as a point on the map, Andrienko et al. [104] proposed a group space in which the positions of individuals are converted from geographical space to an abstract space. Individuals are marked as points in the group space to analyze relative locations of individuals with respect to group movement.

The advantage of point-based visualization is that it enables users to clearly observe individual objects or events in the data. But when the number of objects or events becomes large, severe visual clutter will make the visualization unclear and hard to interpret. Heatmaps (Fig. 5a) with

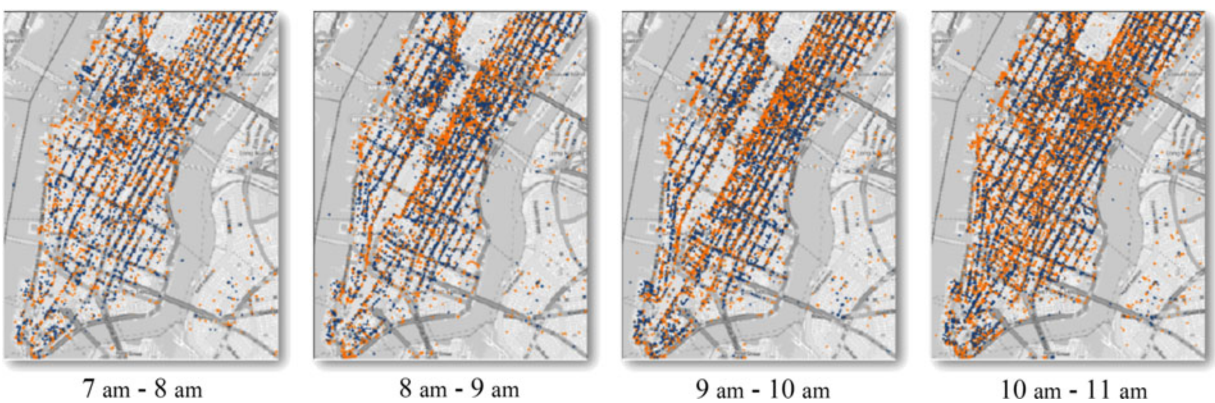
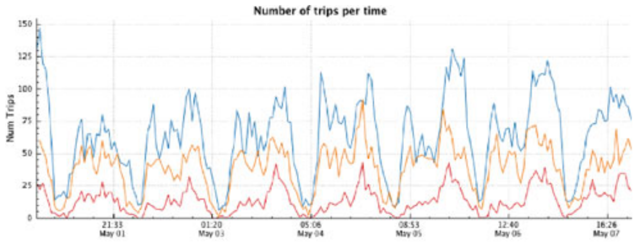
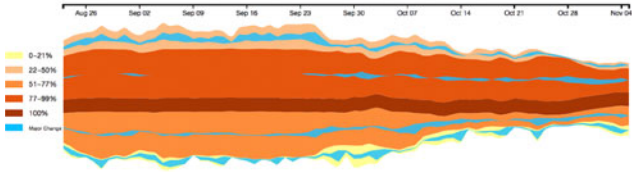


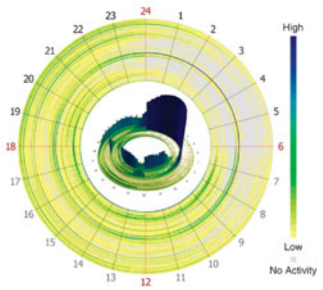
Fig. 2. Example of point-based visualization of locations: Pickups (blue) and drop-offs (orange) of taxi trips in Manhattan from 7 to 11am on May 1, 2011 are labeled by colored points. Notice that during 8-10 am, there are virtually no taxi trips along 6th Avenue, implying the traffic was blocked because of Five Boro Bike Tour [24].



(a) Using *line charts* to represent linear time [24]. It shows the number of taxi trips originating in three regions (lines in different colors) of New York City from May 1 to May 7, 2011.



(b) Using *ThemeRiver*, a well-known type of *stacked graph*, to represent linear time [64]. It shows the evolution of the covered area by the Antarctic ozone hole from Aug. 23 to Nov. 5, 2012.



(c) Adopting *circular time axis* to visualize periodic time [95]. It shows hotness of activity during 24 hours of a day for 96 human activities (one activity per ring), and a complementaty 3D representation is shown in the center to facilitate peak identification.

Fig. 3. Axis-based design for visualization of time.

kernel density estimation (KDE) serve as a common solution to tackle this problem [27], [105], [106].

3.2.2 Region-Based Visualization

Region-based visualization is often used to show aggregated information based on regions of a predetermined division of space. Choropleth map (Fig. 6a) is a typical example of region-based visualization, which shows regions as area marks using given geometry and an attribute is encoded with color [107], [108].

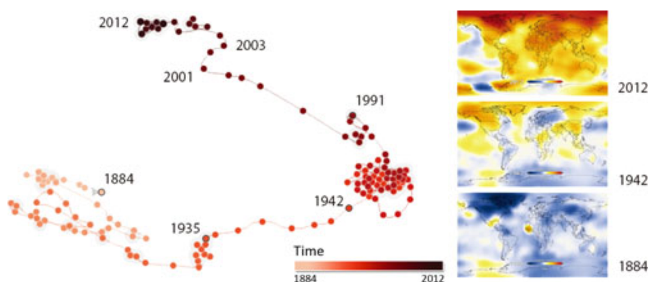
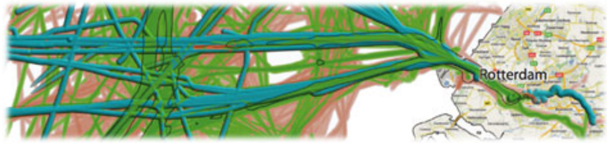


Fig. 4. Using *color* and *connection* to present relative time of data [96]. It shows the evolution of global temperature between 1884 and 2012, with points representing annual temperature world maps of these years and being connected in a chronological order. Color encodes relative time of each data point. Three representative temperature maps for three years are shown (right).

(a)



(b)

Fig. 5. Examples of using heatmap and density map to visualize locations for a large scale of objects, events or trajectories: (a) Visualizing hotspots in a city [27]. The locations with a large number of vehicles passing by are shown in red. (b) Visualizing accident risk based on trajectories of passenger (turquoise), cargo (pink), and tanker (green) vessels in front of Rotterdam harbor [37].

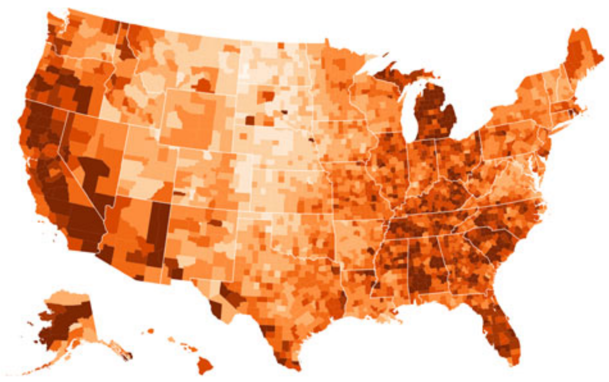
Moreover, to visualize the flows between regions, the flow map can be embedded [89], [109]. However, a flow map easily becomes illegible due to the massive interactions and overlapping of flows as the data size increases. Currently, Guo et al. [110] proposed a new approach to flow smoothing and mapping based on a novel definition of flow neighborhood that addressed three major problems with flow map, including the cluttering problem, the modifiable area unit problem and the normalization problem. One example of smoothed flow map for migration patterns in the USA is shown in Fig. 6b. In addition to the flow map, Zeng et al. [14] designed a circos figure to visualize interchange patterns among different regions of a city (Fig. 6c).

In general, region-based visualization has advantages in revealing macro patterns (e.g., flows among regions), while inadequate for analyzing micro patterns (e.g., individual's behavior) [11]. Therefore, this type of techniques is often used in combination with other techniques to support a comprehensive analysis with different levels of detail.

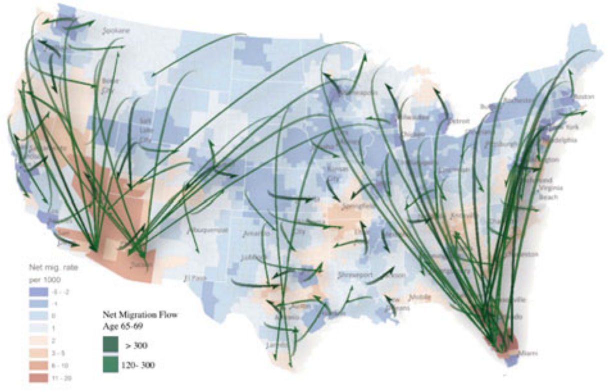
3.2.3 Line-Based Visualization

In urban context, it is common to specify locations based on road maps or traffic networks (i.e., linear referencing, refer to Section 2.2). Line-based visualization can be used to present such type of locations [1], [13], [26], [27], [57], [112]. An example is given in Fig. 7a.

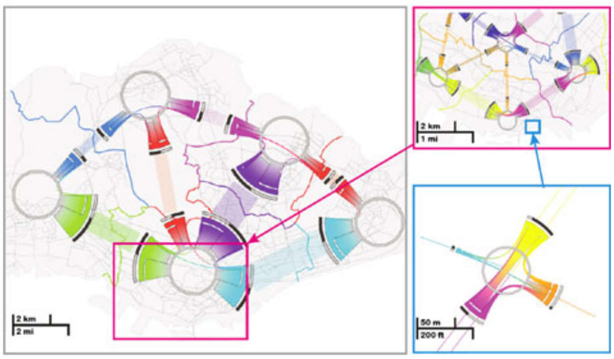
In addition, with the improvement of positioning technologies, it becomes possible to turn the discrete data points into a continuous form, called trajectories, which can tell more semantic information in the space. Line-based visualization can also be used to depict locations on the basis of trajectories. Conventionally, a trajectory can be drawn as a line or curve on a map connecting from the initial point to the last point sequentially [33]. To facilitate the reveal of hidden patterns, trajectories can also be transformed and presented in other spaces using topological or geometric algorithms [104], [113], [114]. For instance, Crnovrsanin et al. [113]



(a)



(b)



(c)

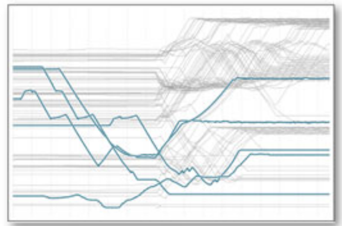
Fig. 6. Examples of region-based visualization: (a) Choropleth map showing US unemployment rates of different regions from 2008, where a darker color means a higher unemployment rate [111]. (b) Visualizing migration flows between different regions for age 65-69 in the USA [110]. (c) The interchange circos figure designed to visualize interchange patterns among regions in city scale (left), regional scale (top right) and road network scale (bottom right) [14].

proposed the proximity-based visualization, transforming trajectories into an abstract space based on proximity data which is computed as the distance between moving objects and some reference locations. Figs. 7b and 7c show an example of applying traditional line-based visualization and proximity-based visualization respectively to visualize people's moving trajectories during the simulation of an evacuation in an office building [113].

Furthermore, for trajectories of large scale, many approaches have been developed to avoid severe visual clutter. First, edge bundling [115], [116], [117], [118] is one of the most popular approaches which groups similar edges into bundles. An example of applying edge bundling on

(a)

(b)



(c)

Fig. 7. Examples of line-based visualization: (a) Visualizing traffic patterns at different locations based on the road network of Beijing [1]. (b) Traditional line-based visualization to show people's moving trajectories during the simulation of an evacuation in an office building after the detonation of an explosive in one of the rooms [113]. (c) Proximity-based visualization which transforms original trajectories into an abstract space where trajectories are plotted as distance to the explosion (*y*-axis) versus time (*x*-axis). After transformation, we can clearly notice the motion of some people (green lines) before most of the other people (grey lines), suggesting possible suspects or witnesses of the event [113].

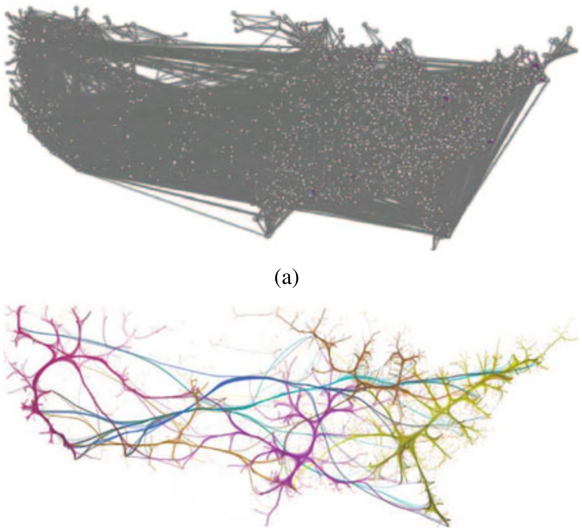
line-based visualization is shown in Fig. 8. Although edge bundling is effective in reducing visual clutter, it introduces visual ambiguities that can impede the understanding of trajectories. In contrast with edge bundling, KDE can also be applied to generate density maps [37], [119] for visualization of a large number of trajectories without twisting them, where density of trajectories are encoded by colors (Fig. 5b).

3.3 Visualization of Other Properties

As shown in Table 1, urban data contains various properties in addition to spatial and temporal information. These properties can be categorized into three types, including numerical properties, categorical properties and textual properties.

Numerical properties refer to those measurements of magnitude that support arithmetic comparison, while categorical properties are discrete attributes that can distinguish whether two data objects are the same or different. Studies have been done to explore proper visual channels to encode numerical and categorical properties. As pointed out by Munzner et al. [108], the same data attribute encoded with two different visual channels will result in different information content through our perceptual system. Therefore, it is necessary to guide the design for the visualization of urban data based on the effectiveness rankings of visual channels for numerical and categorical properties as summarized in [108].

In addition, textual properties refer to information that are recorded by text, such as content of posts on social network, name of road segments and clinical history. These



(a)

(b)

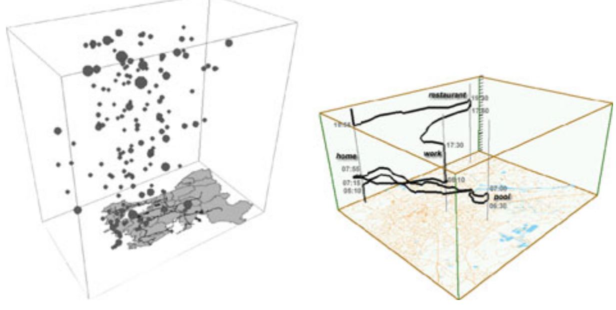
Fig. 8. Examples of edge bundling for line-based visualization of US migrations [120]. (a) Original line-based visualization. (b) Applying edge bundling to reduce visual clutter.

properties provide rich semantic information which is essential for in-depth analysis and interpretation. Text-based visualization techniques, such as Wordle [121], [122], [123], can be employed for visualization of textual properties. Meanwhile, nature language processing (NLP) techniques can also be integrated to extract information and transform textual properties into other types [124].

3.4 Visualization of Multiple Properties

3.4.1 Spatio-Temporal Visualization

When analyzing urban data for real-world applications, it is often necessary to consider both spatial and temporal properties simultaneously. Space-time cube is a typical technique for spatio-temporal visualization [125]. It follows the idea of mapping two spatial dimensions to two axes (i.e., x-axis and y-axis) of a virtual three-dimension cube and use the third dimension (i.e., z-axis) for the mapping of time. The spatial context is often depicted by a map that constitutes one face of the space-time cube. Thus, one can place graphical objects in the cube in order to mark events of



(a) (b)

Fig. 9. Space-time cube for spatio-temporal visualization, which presents spatial dimensions along x-axis and y-axis, and time along z-axis. (a) Visualizing earthquake events in a space-time cube [126]. (b) Visualizing the trajectory of a person's movement in a space-time cube [125].

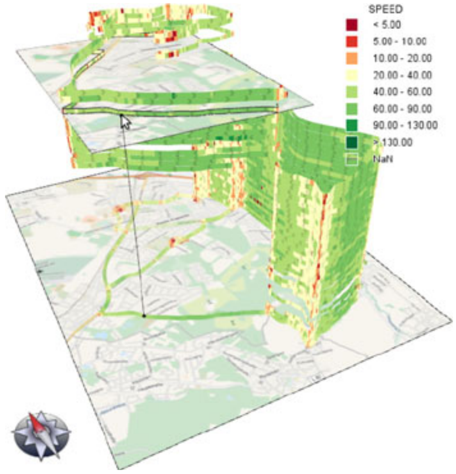


Fig. 10. A variant of space-time cube that visualizes trajectories as stacked 3D bands along which color encodes velocity [29].

interest (Fig. 9a) [126], [127], [128], or construct trajectories that illustrate path of objects (Fig. 9b) [89], [125]. To fulfill the needs of real-world application, the standard space-time cube can be enhanced to depict associated attributes [29], [129]. A typical example is given in Fig. 10, which visualizes a series of trajectories in a space-time cube as stacked 3D bands along which velocity is encoded by color.

However, the use of 3D visualization is controversial, and may lead to ambiguity due to perceptual problems like potential occlusions [108], [130]. Therefore, for an effective analysis, space-time cube usually relies on appropriate interactions to allow users to view the data from different perspectives.

In addition, small multiples [131] and value flow map [132], [133] can also be used to visualize spatio-temporal data.

3.4.2 Multivariate Visualization

Besides spatial and temporal information, many other properties are also involved when analyzing urban data. For most of the time, those properties are correlated, and analyzing such correlations is essential for real world applications. Therefore, visualizing multiple properties simultaneously is required. Compared with multiple coordinated views, multivariate visualization techniques can be employed to visualize urban data in a more compact way, which is particularly important for data of high-dimensions. Recently, Liu et al. [134] presented a survey on advances of multivariate visualization in the past decade. In this section, we mainly focus on visualization techniques which are widely used in urban context. In addition to the conventional 2D/3D charts, these techniques are summarized into four categories (i.e., pixel-based techniques, geometric projections, icon-based techniques, and hierarchical techniques) following the traditional taxonomy by Keim et al. [135].

- **Pixel-based Techniques**

Pixel-based techniques are a popular type of visualization for urban data. The basic idea is to map data values to pixels with a proper color scheme and then group the pixels adequately for analytical tasks. This type of techniques enjoys better scalability for displaying data items in a single view than other techniques. For existing work in urban visual analytics, matrix is the most common form [1], [136], [137], [138], to arrange pixels due to its interpretability and simplicity. Moreover, the circle segment is a variant of the matrix form, whose idea is to represent data in a circle which is divided into segments, one for each attribute. The ring map [95] is an example of this type (Fig. 3c).

Although pixel-based techniques enjoy relatively better scalability, human performance with this type of visualization may be influenced by a number of factors, such as screen resolution, working memory and attention demands for a given task. Meanwhile, as data is so densely rendered in the view, it leads to a heavy cost for online interactions and animated transitions.

- **Geometric Projections**

Geometric projections aim to find informative projections and transformations for multivariate data so as to use the spatial position channel to visually encode them [135]. Due to the complexity and homogeneity of urban data, finding an appropriate projection is not an easy task. Therefore, the use of methods in this category for urban visual analytics is quite limited, mainly based on two traditional approaches, scatterplots [1], [64] and parallel coordinates [17], [139].

In a *scatterplot*, data items are mapped along x and y axes of a Cartesian coordinate system defined by two data properties or multidimensional projections like principle component analysis (PCA) or multidimensional scaling (MDS). Scatterplot matrix is a variant of scatterplot in which multiple pairwise projections of data properties can be shown simultaneously. However, the scalability of a scatterplot matrix is limited to around one dozen attributes and hundreds of items [108], and will suffer a scalability problem when the dimensionality or data size becomes larger.

A *parallel coordinates plot (PCP)* presents a multivariate data item as a polyline with vertices on parallel axes, and each axis corresponds to one data property. However, parallel coordinates suffer from two major limitations. One is how to determine the order of axes, as different axis order will highlight different aspects of the data structure. The other is the visual clutter problem caused by a large number of polylines.

- **Icon-based Techniques**

Icon-based techniques aim to provide a compact form to map properties of each multivariate data item to graphical features of an icon or glyph. For urban visual analytics, there are many examples of glyph usage, especially in conjunction with other visualizations, such as maps [140], scatterplots [64] timelines [141] and graphs [45], to facilitate a comprehensive analysis. It is widely believed that a good glyph design can facilitate effective learning, memorizing and comparing, while a less effective glyph design may suffer from perceptually confusing, semantically ambiguous, or hard to learn and remember [142]. In general, it is not an easy task to design an effective glyph. Even though most existing designs have undergone an enduring process of evolution, refinement and standardization [142], one of the most common criticisms of using glyphs is that there is an implicit bias in most mappings (i.e., some graphical features or relationships between features are easier to perceive than others) [143].

- **Hierarchical Techniques**

Hierarchical techniques are mainly designed to visualize the hierarchy structure of data, including both dimension hierarchies [17], [144], [145] and data hierarchies [20], [22] [146], [147]. It splits the data space into subspaces and organizes them in a hierarchical manner. For visualization of urban data, node-link diagram and treemap are two frequently used visual forms. In a *node-link diagram*, data items are presented as nodes with links among nodes representing corresponding relationships. There are various applications of node-link diagrams in urban context including community visualization [20], [148], topic analysis [146], and information diffusion interpretation [22], [149]. *Treemap* is a space-filling method that partitions the visual space into regions and shows hierarchical relationships with containment. Based on the classic treemap, many extensions, such as Voronoi treemap [150], Nmap [151] and contour-based treemap [17] have been proposed for different analytical tasks and application scenarios.

4 Combination of Visualization and Automated Analytical Approaches

Automated data analysis techniques enable convenient exploration of data and discovery of patterns, such as clusters, trends and anomalies. However, due to the growing mass and increasing complexity of urban data, automated approaches suffer more and more challenges. In the mean time, some information is hard to be precisely quantified in certain applications (e.g., personalized preference), which makes automated approaches even not applicable. On the other hand, visualization has now become a standard technique that can help involve human perceptual capabilities into the automated data exploration process. Through the combination of visualization and automated analytical approaches, high-level and complex tasks can be performed more effectively and efficiently.

In this section, we discuss how visualization can be combined with automated analytical approaches, which is one of the core issues of urban visual analytics. Based on the existing work, we believe the benefit of this combination is not limited to model visualization and building as described in the general frame work of visual analytics by Keim et al. [152]. It can bring various benefits and has great potential for future development. For example, visualization can be adopted to provide different levels of data summarization for the hypothesis investigation, and facilitate a proper choice of automated approaches for further analysis and mining. Results generated by automated analytical approaches can also be presented via certain visual forms to achieve a better understanding and interpretability. Meanwhile, visualization can be integrated into an automated machine learning process to make it more effective and flexible for various applications. Therefore, we survey the existing urban visual analytics tools and categorize them into two classes based on their different outputs (Fig. 11), which also implies different roles that visualization plays. One class is "*data exploration and pattern interpretation*" for which visualization enables analysts to explore data interactively and gain better insights of existing patterns detected by automated approaches. The other class is "*visual learning*"

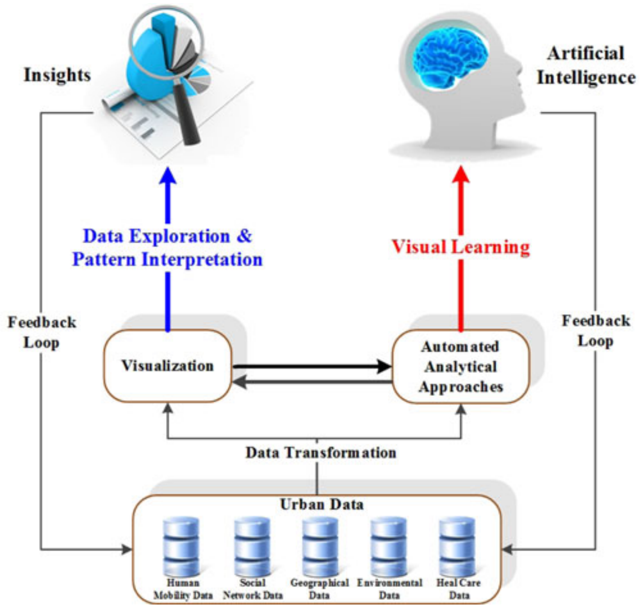


Fig. 11. Two types of combination of visualization and automated analytical approaches for existing urban visual analytics tools.

where analysts seek the help of visualization to guide automated data mining or machine learning process, leading to sort of artificial intelligence, such as a classifier [40], a predictor [53] or a regression model [67], which can be applied on a larger scale of data for further exploration.

4.1 Data Exploration and Pattern Interpretation

In recent years, a large number of urban visual analytics tools have been developed, which combines visualization with automated analytical approaches for data exploration and pattern interpretation in urban context. They cover different urban data types, including human mobility data, social network data, and environmental data. In the following part, we present representative works for different data types and summarize them in Table 2.

4.1.1 Human Mobility Data

Human mobility data is one of the most popular types of urban data. Lots of work has been done for various applications, such as pattern interpretation [1], [14], [17], [26], [27], [106], understanding of movement [12], [153], [154], [155], monitoring of transportation system [13], [16], [30], [38], [58], [156], [157], and incident detection [15], [19], [158].

First of all, many studies focus on interactive methods to detect and explore various patterns hidden in human mobility data. For instance, Wang et al. [1] presented a visual analytics system for the analysis of patterns in traffic congestions. Traffic congestions are characterized by low speed on the road. The system not only enables the road segment level analysis, but also provides a comprehensive exploration on the propagation graph level which depicts the propagation of a traffic congestion in time and space. TelCoVis [17] (Fig. 12) is a visual analytics system for the co-occurrence pattern in urban human mobility, a pattern of high social and business values. A series of coordinated visualizations are provided to gain insights of co-occurrence and analyze correlations of co-occurrence events through biclusters. Liu et al. [27] proposed a visual analytics approach to explore route diversity (Fig. 13b). The system supports a multi-scale analysis of diversities which can help reveal the importance of different road for different trips and support urban planning and management. In addition, Zeng et al. [14] looked into interchange patterns aiming at revealing passenger redistribution in a transportation network (Fig. 13a), while Zheng et al. [106] studied bi-directional movement patterns which exist ubiquitously in our daily life. Moreover, Huang et al. [26] proposed TrajGraph, a new visual analytics method, for studying urban network centralities by integrating graph modeling with visual analysis based on taxi trajectory data. *LIVE Singapore!* [156] is another representative project by MIT Senseable City Lab which developed interactive applications (Fig. 14) enabling experts and citizens to gain a better understanding on how Singaporeans move through urban space and explore the various narratives found within urban mobility.

Besides pattern analysis, some researchers tried to develop visual analytics tools to facilitate understanding of movement in urban context. Andrienko et al. [153] investigated the aggregation methods for visual analysis of movement data from two aspects, traffic-oriented and trajectory-oriented, to support analytical tasks for city traffic management. Poco et al. [159] explored traffic dynamics in urban environments using vector-valued functions. Further, Von et al. [154] proposed MobilityGraphs, a graph based method, to reveal the variation of the presence of people in different places over time as well as the movement flows between places. Clustering is utilized to support data exploration from different aspects and at different scales. Moreover, Wang et al. [12] presented a visual analytics system to understand urban traffic based on sparse traffic trajectory data (i.e., loop sensors data) through global exploration, cell exploration and correlation exploration (Fig. 15). Local animation and aggregation methods are utilized to address the uncertainty issue.

In addition, transportation is a hot topic in urban computing, thus recently many visual analytics tools are developed to monitor the transportation system based on human mobility data. Zeng et al. [13] explored passengers' mobility in the public transportation system of Singapore based on passenger RFID card data, enabling visualization and exploration of various mobility related factors such as riding time, transfer time and waiting time (Fig. 16). Palomo et al. [58] proposed TR-EX, a visual analytics system, to study the New York City subway service based on transportation schedules. Moreover, Lorenzo et al. [16] presented AllAboard, an intelligent tool to enable city planners visually explore urban mobility and interactively optimize public transportation based on mobile phone data. There is also work based on surveillance videos, such as analyzing the movement of recorded objects [157] and monitoring traffic conditions in a tunnel [38].

Other work based on human mobility data is dedicated to detecting and investigating incidents in urban context. Andrienko et al. [158] came up with a visual analytics procedure for analyzing place-related events. The procedure contains four steps: 1) event extraction from trajectories, 2) determine relevant places, 3) aggregating events and trajectories, and 4) analysis of the aggregated data. Meanwhile, Andrienko et al. [15] also proposed a suite of visual analytics

TABLE 2
Representative Urban Visual Analytics Tools for Data Exploration and Pattern Interpretation

Category	Tools / Systems	Characteristics	Datasets
Human mobility data	Huang et al. [26]	Interpretation of urban transportation patterns	Taxi trajectories in Shenzhen, China
	Wang et al. [1]	Exploration of urban traffic congestions	Taxi trajectories in Beijing, China
	Liu et al. [27]	Exploration of route diversity	Taxi trajectories in Shanghai, China
	Zheng et al. [106]	Interpretation of bi-directional movement patterns	Taxi trajectories in Shanghai, China
	Wang et al. [30]	Analysis of real traffic situations	Taxi trajectories in Hangzhou, China
	Kloeckl et al. [156]	Exploration of mobility patterns in Singapore	Taxi trajectories in Singapore
	Andrienko et al. [158]	Extract and characterize significant places of a city	Car trajectories in Milan
	Andrienko et al. [153]	Spatio-temporal aggregation for movements	Car trajectories in Milan
	Poco et al. [159]	Exploration of traffic dynamics using vector-valued functions	Taxi pickups and dropoffs in New York City
	Palomo et al. [58]	Interpretation of spatio-temporal patterns in transportation services	New York City subway service
	Zeng et al. [13]	Exploration of passenger mobility in a public transportation system	Passenger RFID card data in Singapore
	Zeng et al. [14]	Revealing interchange patterns of passengers in a transportation network	Passenger RFID card data in Singapore
	Wu et al. [17]	Exploration of co-occurrence pattern in urban area	Mobile phone data in Guangzhou, China
	Lorenzo et al. [16]	Exploration of urban mobility for public transport optimization	Mobile phone data in Abidjan
	Andrienko et al. [15]	Reconstructing past events from activity traces	Mobile phone data in Milano; Flickr photos made on British Isles
	Von et al. [154]	Clustering of mass mobility flow via spatio-temporal graphs	Geo-tagged twitter in London; mobile phone data in Abidjan
	MacEachren et al. [19]	Situational awareness of crisis	Geo-tagged twitter
	Wang et al. [12]	Exploration of sparse traffic trajectory data	Traffic loop sensors data in Nanjing, China
	Meghdadi et al. [157]	Visualization of objects' trajectories in videos	Surveillance videos
	Piringer et al. [38]	Situation awareness of road tunnels	Surveillance videos
Social network data	Liu et al. [47]	Exploratory microblog retrieval	Twitter
	Cao et al. [45]	Exploration of anomalous user behaviors	Twitter
	Sun et al. [48]	Interpretation of topic coopetition	Twitter
	Wu et al. [51]	Interpretation of opinion diffusion	Twitter
	Zhao et al. [22]	Interpretation of anomalous information spreading	Twitter
	Zhao et al. [52]	Understanding personal emotion style	Twitter
	Xu et al. [21]	Interpretation of topic competition	Twitter
	Bosch et al. [43]	Real-time monitoring of microblog messages	Twitter
	Cao et al. [44]	Tracing spatio-temporal process of information diffusion	Twitter
	Chae et al. [46]	Interactive abnormal event detection	Twitter
	Krueger et al. [160]	Interactive exploration of movement semantics and handling semantic uncertainties in space and time	Foursquare
	Kwon et al. [161]	Exploration of conversation threads of online health communities	OHC forum data
Environmental data	Accorsi et al. [62]	Exploration of river water quality	River water quality data
	Li et al. [69]	Exploration of climate change data	Climate change data
	Goodwin et al. [65]	Exploration of energy consumption data	Energy consumption data
	Qu et al. [61]	Analysis of the air pollution problem in Hong Kong	Air quality and weather data

Fig. 12. TelCoVis, a visual analytics system to explore co-occurrence pattern in urban human mobility based on mobile phone data [17].

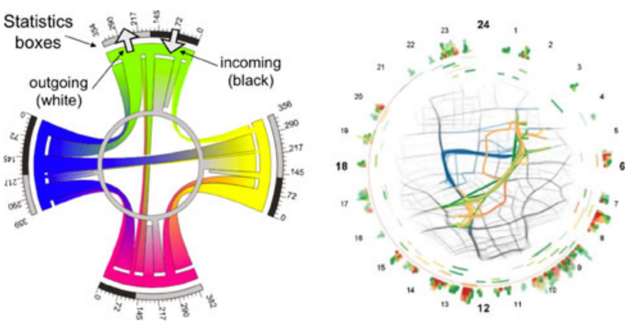
methods for detecting and reconstructing events from people's activity traces. Besides, SensePlace2 [19] is a visual analytics system developed to support situation-aware exploration of crisis events based on geo-tagged twitter.

4.1.2 Social Network Data

Based on social network data, many interesting topics are investigated, among which topic evolution and information diffusion are two popular topics in recent years.

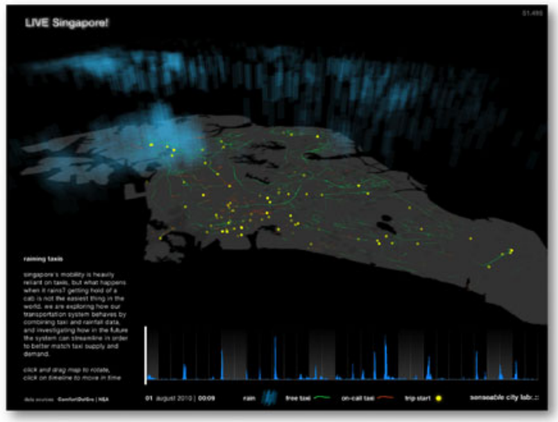
Xu et al. [21] proposed a timeline visualization tool (Fig. 17) to interpret the competition for public attention on multiple topics promoted by various opinion leaders on social media. The tool features both topical and social aspects of the information diffusion process. For the design, ThemeRiver shows the increase and decrease of competitiveness of each topic, while opinion leaders are drawn as threads which converge or diverge with regard to the change of their roles in influencing the public agenda over time. Similar work studying the evolution of topics or opinions on social media includes EvoRiver [48] and OpinionFlow [51].

Cao et al. [44] proposed Whisper, a visual analytics system for tracing of information diffusion process. Following the metaphor of a "sunflower" (Fig. 18), three novel visual ingredients: a hierarchical social-spatial layout, pathways of information flow, and a dynamic diffusion series are designed to characterize different diffusion processes in Twitter, including collective responses, multi-step information flows and temporal heterogeneity. Moreover, Zhao et al. [22] presented #FluxFlow to study how rumors spread in the social network.

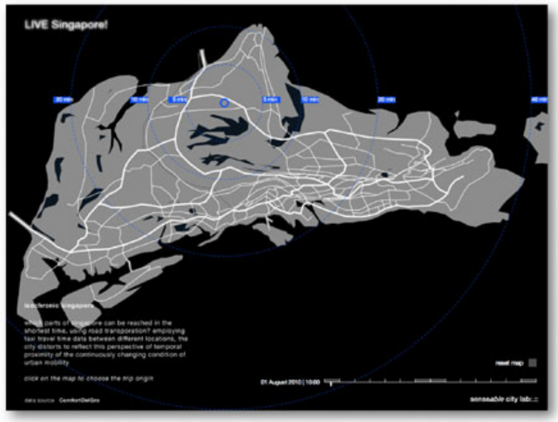


(a) (b)

Fig. 13. (a) Exploration of interchange patterns in massive movement data [14]. (b) Interpretation of route diversity of taxi trips in urban area [27].



(a)



(b)

Fig. 14. Two visualizations of *LIVE Singapore!* [156]. (a) Visualizing taxi trips and demands in Singapore during heavy rainfalls. (b) Distorting Singapore's shape on a map to reflect the time it takes to travel from the selected origin to other parts of the island.

Furthermore, social network data offer a great opportunity for us to understand urban dynamics by providing rich semantic information which is essential for in-depth analysis in real-world applications. Therefore, another interesting research direction should be combining spatial, temporal and textual information extracted from raw social network data to support interactive exploration and interpretation on urban issues. For example, VAST Challenge 2011 [162] provided a typical scenario in this direction, where comprehensive analyses of social media reporting illnesses in a city are conducted to support decision making on a public health crisis. Nabian et al. [163] proposed the design of the MIT GEOblog platform, allowing people to share contents based on their real-time locations sensed by the system. Krueger et al. [160] presented a visual analytics approach to enrich movement data with POI information using social media services and handle semantic uncertainties in time and space via a POI decision model in combination with highly interactive visualizations (Fig. 19).

In addition, there is also work based on social network data studying other topics, including anomalous user detection [45], event investigation [43], [46], conversation threads exploration [161], emotion analysis [52], [164] and explorative microblog retrieval [47].

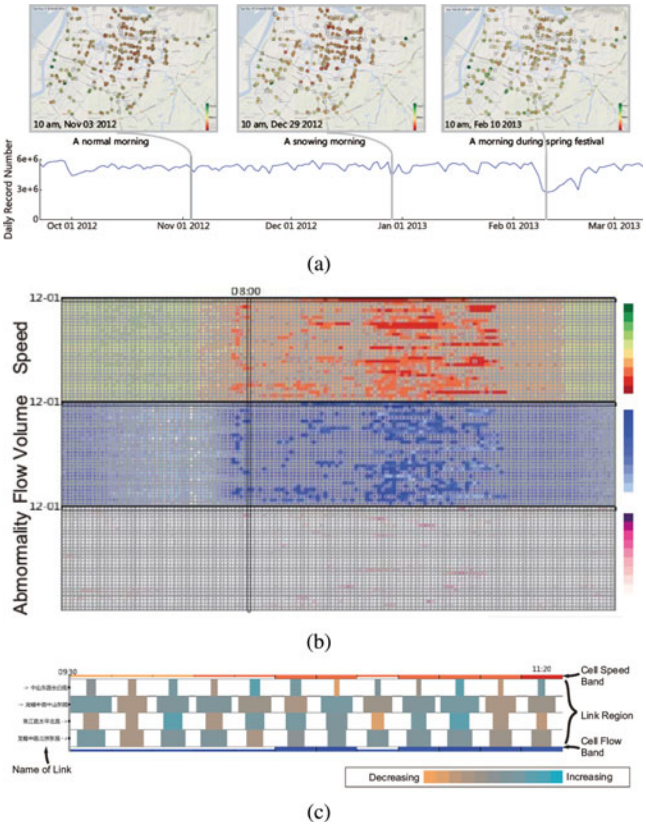


Fig. 15. A visual analytics system [12] for understanding of urban traffic based on sparse traffic trajectory data through (a) global exploration, (b) cell exploration and (c) correlation exploration.

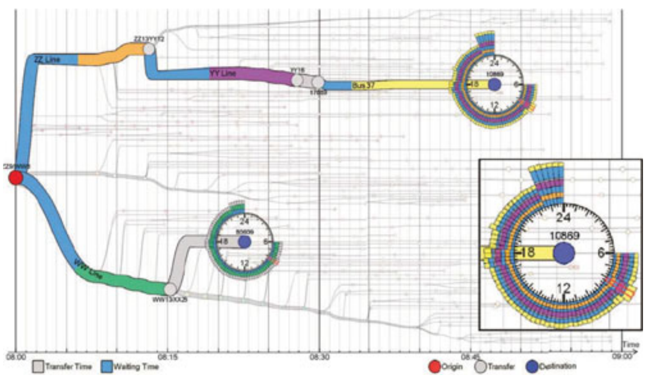


Fig. 16. Visual exploration of mobility of public transportation system in Singapore based on RFID card data [13].

4.1.3 Environmental Data

Environment is an important topic in urban computing. In the field of visualization, some work has been conducted to explore different environmental datasets. Qu et al. [61] presented a visual analytics system to study the air pollution problem in Hong Kong. Several novel visualizations, including polar systems embedded with circular pixel bar charts, enhanced parallel coordinates with S-shape axis, and weighted complete graphs, are proposed to support multiple analytical tasks, such as attributes correlation detection, data comparison and air quality trend identification. Other related work includes analysis of river water quality [62], climate change [69], [165], [166], and energy consumption [65].

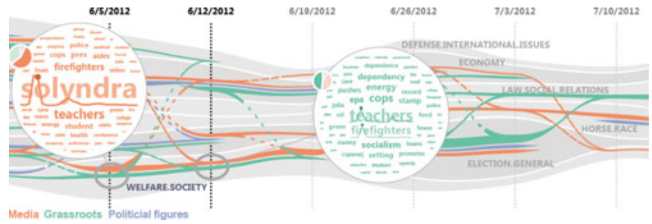


Fig. 17. Interpretation of topic competition on social media [21].

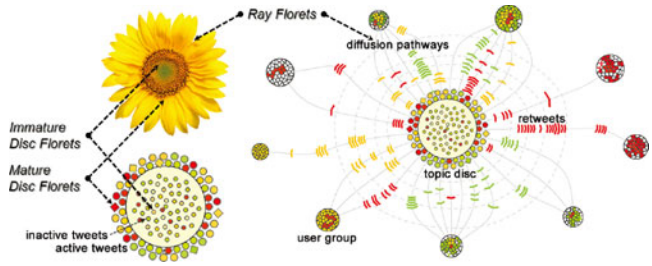


Fig. 18. Whisper: a visual analytics system with the metaphor of sunflower for tracing information diffusion process in Twitter [44].

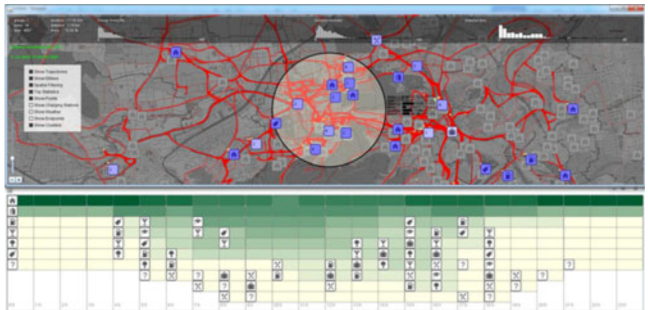


Fig. 19. A system for interactive analysis of semantically enriched movement data with POI information using social media services and handling semantic uncertainties in time and space [160].

4.2 Visual Learning

The key objective of visual analytics as well as its challenge is to identify the limits of analytical algorithms which can not be appropriately automated, and then develop a tightly integrated solution that adequately integrates automated approaches with appropriate visualization and interaction techniques [167]. *Visual learning* is an effective way to achieve this objective. It is a special case of active learning [168] in which an automated machine learning or data mining algorithm is able to interactively query users through visualizations to optimize and obtain desired outputs. Here we first discuss how visualization can help during three major steps of automated analytical approaches, including 1) cohort construction; 2) feature selection and model construction; 3) result evaluation and tuning, then present the representative applications in urban visual analytics and summarize them in Table 3.

4.2.1 Cohort Construction

Cohort construction is usually the initial stage of an automated analytical approach. In this stage, analysts need to choose proper data samples from a labelled dataset as the input of a machine learning or data mining algorithm. Ideally, samples in the dataset are assumed to be well labelled.

TABLE 3
Representative Urban Visual Analytics Works for Visual Learning

Works	Characteristics	Steps Adopting Visualization		
		Cohort Construction	Feature & Model Selection	Result Evaluation & Tuning
Andrienko et al. [23]	Interactive visual clustering of large collections of trajectories	√		√
Muhlbacher et al. [67]	Regression model construction		√	
Lu et al. [53]	Developing predictive models utilizing social media data		√	√
Poco et al. [138]	Visual reconciliation technique for model comparison in climate science		√	√
Arietta et al. [173]	Identifying and validating predictive relationships between visual appearance of a city and its non-visual attributes		√	√
Chen et al. [18]	Tackling sparsity problems of geo-tagged social network data for exploration of movement patterns	√	√	
Quinan et al. [174]	Examining behaviors of and relationships among weather features		√	
Yu et al. [40]	Classification for the home-work pattern			√
Liao et al. [175]	Detecting anomalies in GPS data			√
Wang et al. [176]	Construction of causal relations			√
Cao et al. [177]	Analysis of multidimensional clusters			√

However, in practice, labels may not exist or lack accuracy. One solution is to label all samples manually, but it takes tremendous efforts to label a large dataset. Besides, an input with more data samples usually means a longer processing time and more potential noises. Therefore, in many scenarios [18], [23], [168], an active cohort construction is needed, whose basic idea is to allow an automated algorithm to choose an input with fewer efforts of labelling but leading to a higher accuracy.

Fig. 20 illustrates this basic idea with an example of binary classification. As shown in Fig. 20a, the dataset contains 400 samples from two Gaussian distributions which are plotted in a 2D space. If we randomly choose 30 samples and label them, the accuracy of classifier trained on the labelled samples is relatively low (i.e., 70 percent accuracy in Fig. 20b). On contrary, labelling 30 carefully selected samples can dramatically increase the accuracy (i.e., 90 percent accuracy in Fig. 20c). The reason is that a few training samples close to the decision boundary are crucial to the accuracy of classification.

Based on the above discussion, we can see that one of the key issues for an effective automated analytical process is the capability of locating the most informative data samples as input (i.e., cohort construction). To achieve this goal, visualization can enable analysts to obtain an overview of a large number of data samples with different levels of detail and provide visual cues to support the labelling process. Thus analysts can choose samples wisely and make the analytical process more effective and efficient.

4.2.2 Feature Selection and Model Construction

Proper feature selection and model construction is a fundamental step for various data analytical tasks. Although many automated computational approaches have been proposed, they may not be efficient due to the large volume of data. Moreover, in the era of big data, as models become more and more complicated, most people, especially domain experts, have difficulties in understanding the algorithm and reflecting their domain knowledge throughout the analysis process. For example, for artificial neural networks (ANNs), a popular computational model in machine learning, it is an indeed difficult, very time-consuming, and highly experience-dependent process to figure out the details of model construction so as to improve the model for a more satisfactory output. Even nowadays, there are rarely efficient ways but to manually adjust the number of layers or neurons within each layer, modify the drop rate or learning rate of the algorithm, or try all potential functions one by one on neurons [170]. In order to tackle this limitation, one possible way is to involve analysts in managing the outcome from each iteration of automated optimizing process through interactive visualization. In this way, analysts can apply their domain knowledge to tune the data model, or adjust the input data features, thus intuitively control and improve the whole process of data exploration. Therefore, in brief, visualization can clearly depict detailed data distribution, help users define a better data model with appropriate features, and avoid redundancy and misunderstanding of data.

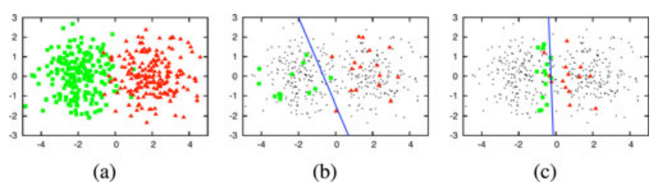


Fig. 20. Illustrative example for the basic idea of an active cohort construction [169]: (a) A dataset of 400 samples from two Gaussian distributions. (b) A classifier based on a logistic regression model trained with 30 randomly labelled samples (70 percent accuracy). (c) A classifier based on a logistic regression model trained with 30 actively queried samples (90 percent accuracy).

4.2.3 Result Evaluation and Tuning

Evaluation of results is also an important issue in data analysis. Automated methods generate the optimal result according to certain pre-defined measurements. However, appropriate measurements are always difficult to find. In another word, as each measurement has its own advantages and disadvantages, analysts often have difficulties in choosing a proper one for a certain analytical task. For example, clustering is a powerful technique to automatically partition data. However, it is difficult to evaluate the quality of clustering results, especially for multidimensional data. Although several measurements, such as Davies-Bouldin index [171], $index - \Im$ [172], have been proposed to compare how well different clustering algorithms perform on a data set, each measurement has its own drawbacks.

To tackle this challenge, when obtaining the results from automated approaches, visualization can display the results directly to analysts who can judge the performance and tune the results based on their domain knowledge, which to some extent, is much more effective and efficient than statistical measurements, especially in practical applications.

4.2.4 Applications in Urban Visual Analytics

As an emerging field, to the best of our knowledge, the applications of visual learning in urban computing are quite limited. In this section, we briefly review some current representative works and summarize them in Table 3.

Andrienko et al. [23] proposed an approach for interactive cluster analysis of a large number of trajectories of moving entities, which are structurally complex. The approach mainly contains four steps. First, the analyst chooses a manageable subset of objects (i.e., cohort construction) and applies clustering to it. And then users can inspect the result and refine the clustering to gain meaningful results with respect to the analytical tasks (i.e, result evaluation and tuning). After that, the analyst can build a classifier based on the clustering result and may also modify the clusters for better conformance to the goals. And finally the generated classifier is applied to the whole dataset. When necessary, the analyst may repeat the procedure (take a subset - cluster - build a classifier - classify) iteratively for unclassified trajectories.

Muhlbacher et al. [67] proposed a visual analytics framework for building regression models to address limitations of automated approaches. The limitations include selecting input variables, identification of local structures, transformations, and interactions between variables. The framework combines a qualitative analysis of relationship structures by visualization (Fig. 21) and a quantification of relevance for ranking any number of features and pairs of features which may be categorical or continuous. A central aspect is the local approximation of the conditional target distribution by partitioning 1D and 2D feature domains into disjoint regions, which enables a visual investigation of local patterns and largely avoids structural assumptions for the quantitative ranking.

Lu et al. [53] presented a framework for the development of predictive models utilizing social network data. Feature selection mechanisms, similarity comparisons and model cross-validations are combined through a variety of

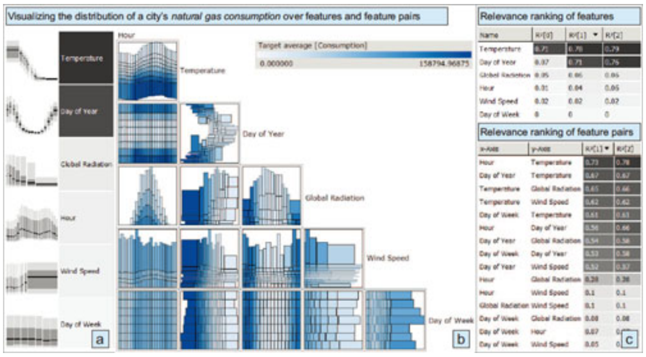


Fig. 21. A visual analytics system for building regression models of natural gas consumption [67].

interactive visualizations to support analysts in model building and prediction. An overview is designed for quick trend analysis with detailed views for tweet sentiment exploration. A similarity widget is embedded to enable analysts to quickly evaluate and compare the accuracy of predictions based on various criteria of similarity, and to perceive the quality of the generated prediction model. Meanwhile, the core component of this framework is an iterative feature selection and model construction module for analysis and comparison.

When analyzing urban data, it often requires grouping of data objects based on their similarity. While extracting groups using a single similarity criteria is relatively straightforward, comparing alternative criteria poses additional challenges. Poco et al. [138] proposed a visual reconciliation technique that helps analysts understand the dependency between alternative similarity spaces for climate models, facilitates iterative refinement of groups, and allows flexible exploration of the parameter space for reconciling the importance of the model parameters with model groups. Similarly, Quinan et al. [174] presented a visualization tool to examine the behaviors of and relationships among weather features. In addition, Arietta et al. [173] introduced a method for interactively identifying and validating predictive relationships between the visual appearance of a city and its non-visual attributes (e.g., crime statistics, housing prices, population density).

Chen et al. [18] proposed a visual analytics system to study sparsely sampled trajectories extracted from geo-tagged social network data which provides rich text and movement information. To tackle the sparsity problem, an uncertainty model based on Gaussian Mixture Model is proposed to characterize the time interval distributions due to various transportation methods. For an effective modeling process, unreasonable time intervals need to be filtered interactively to generate appropriate cohort as an input. In addition, users could also adjust the parameters of models, such as the output confidence interval range and the number of transportation methods (corresponding to the number of gaussian kernels in the model) for a better match based on prior knowledge. Thus, through proper cohort construction and model selection, users can effectively model the movement patterns and explore the semantics based on sparse geo-tagged social network data

Yu et al. [40] presented iVizTRANS, a tool which combines an interactive visual analytics component to aid urban

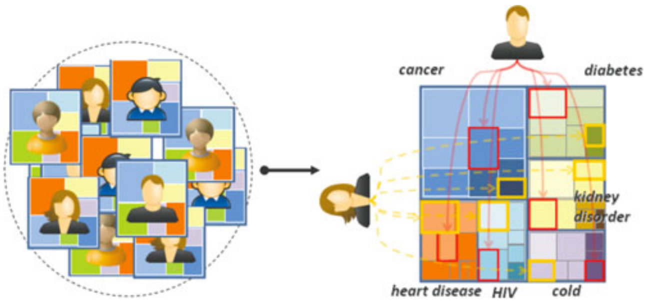


Fig. 22. DICON: evaluation and tuning of multidimensional clustering results [177].

planners to analyze complex travel patterns and decipher activity locations for single public transport commuters. It is coupled with a machine learning component that iteratively learns from the planners' adjustment of classifications (i.e., result tuning) to train a classifier which is then applied to the city-wide smart card data to derive the dynamics for all public transport commuters.

Liao et al. [175] introduced GPSvas, a visual analytics system that detects anomalies in GPS data. In the system, a conditional random field (CRF) model is used as the machine learning component for anomaly detection in streaming GPS traces. Meanwhile, a visualization component and an interactive user interface are built to visualize the data stream, display significant analysis results (i.e., anomalies or uncertain predications) and hidden information extracted by the anomaly detection model, which enables human experts to observe the real-time data behavior and gain insights into the data flow. In addition, analysts can choose to browse the most relevant information to further provide guidance to the machine learning model through interaction and the learning model is then incrementally improved.

Wang et al. [176] presented a visual causal analyst, a novel visual causal reasoning framework that allows users to apply their expertise, verify and edit causal links, and collaborate with the causal discovery algorithm to identify a valid causal network. Its interface consists of both an interactive 2D graph view and a numerical presentation of salient statistical parameters, such as regression coefficients, p-values, and others. Both help users in gaining a good understanding of the landscape of causal structures particularly when the number of variables is large.

Cao et al. [177] designed DICON system which uses dynamic icons (Fig. 22) to represent a multidimensional cluster. The quality of clusters can be conveniently evaluated with the embedded statistical information. And through rich interactions (i.e., merge, split, filter, regroup data within clusters), analysts can further refine clustering results more efficiently.

To summarize, although automated approaches have achieved some successes over past decades, limitations become more obvious with the coming era of big data. The combination of visualization and automated analytical approaches provides us with a potential solution to tackle these limitations and accomplish challenging tasks. The need of such type of combination becomes more and more intense with the increasing complexity of data and analytical tasks, especially in a multi-disciplinary field like urban computing, where advanced data science meets conventional disciplines. Although some initial work has been done in recent years, more efforts are needed to establish an effective framework for this type of combination for various real world applications.

5 Conclusion

Undoubtedly, we are in the midst of a data explosion. The social and economic potential of data is widely recognized. This presents both unprecedented opportunities and challenges to us. On one hand, there are reasons to believe that more and more data will become accessible, which offers precious information resources and opportunities to study and build a better world. On the other hand, although research in urban visual analytics has been hot for the last few years, there are still quite a few challenging issues, such as scalability, heterogeneity, sparsity and uncertainty, that have not been addressed satisfactorily, while new research topics keep emerging due to the boom in big data analytics. Based on our survey, we point out several potential research directions worth further study as follows:

First, with the improvement of data acquisition techniques, data of new types keep emerging, most of which cannot be visualized using existing methods directly. On the other hand, the sparsity and uncertainty of data become more and more remarkable. Pre-processing of these raw data in order to use it for visual analysis bears several potential quality problems. Data can be inherently incomplete or imprecise due to sampling errors or fuzziness caused by privacy protections. Then how to generate an appropriate visual design insensitive to data quality issues, or how to explicitly visualize errors and uncertainties in data to make the analysts aware of the problem poses great challenges. Although some initial work [18], [178] has been carried out, methods are far from enough to deal with these challenges effectively. Thus, how to visualize these new types of data with consideration of data sparsity and uncertainty needs to be further studied in urban visual analytics.

Second, with the increasing complexity of analytical tasks, urban big data analysis often requires synthesis of heterogeneous types of data [179]. So far, it is not an easy task to investigate the implicit relationship among multiple data sources, and current visual analytics tools usually only support single data type. Even worse, some applications may also require visualization of streaming data to support real-time decision making within situation-aware and immersive environments. In order to meet the challenging analytical requirements in the foreseeable future, new techniques for urban visual analytics capable of efficiently handling heterogeneous data and streaming data will be needed.

Third, scalability in general is a key challenge of visual analytics, especially for dealing with the wide variety of big data collected in urban space. As the size of data is continuously growing in terms of both the number and dimension of data items, the compression rate to visualize the dataset keep increasing, and therefore, more and more details are lost. It is a future task of urban visual analytics to create a high-level view of these urban big data to gain insight, while maximizing the amount of details at the same time.

Fourth, we deem that most urban visual analytics applications should be developed using a user-centered design approach [108] and evaluation should take place in several phases along the process. For a complete evaluation, different aspects, including user preference, insight generation, task performance and algorithmic efficiency, should be considered systematically. A comprehensive comparison with existing techniques or tools may also be required to assess the adequacy. In this context, objective rules of thumbs to facilitate design decisions on urban visual analytics would be important contributions to the field.

Moreover, we believe an effective way to mine treasures from urban big data is to let more and more people not only play the role of data consumers, but participate in data analysis process through crowdsoursing. The term "crowdsourcing" was first coined in 2005 by the Wired Magazine [180], which is defined as the practice of obtaining needed services or content by soliciting contributions from a large group of people. Visualization, as a technique that harnesses human perception and cognition capabilities to fulfill tasks, can offer remarkable opportunities in facilitating crowdsourcing to tackle challenges in urban computing. Despite some initial attempts [181], [182], it will be an interesting direction with real potential to explore the benefits of combining visualization and crowdsourcing.

Finally, it is widely accepted that human perception plays an essential role in data analysis. With the increasing volume and complexity of urban data, it is unlikely that computer can take the place of human beings in a decision-making process in the foreseeable future. It will be more and more important to figure out an efficient way to involve humans in the process of data analysis. Although visualization provides a potential solution, it should not stand alone, but should integrate seamlessly into the applications of diverse domains. Thus, how to combine the advantages of visualization and automated analytical approaches to establish more efficient methods for data analysis is crucial and will be a hot subject for future research.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments. This research was supported in part by HK RGC GRF 618313, the National Basic Research Program of China (973 Program) under Grant No. 2014CB340304, and University of Macau Grant SRG2015-00050-FST. Wenchao Wu is the corresponding author of this article.

REFERENCES

[1] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. v. d. Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013.

[2] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, 2013, pp. 1436–1444.

[3] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, 2014, Art. no. 38.

[4] E. Paulos and E. Goodman, "The familiar stranger: Anxiety, comfort, and play in public places," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2004, pp. 223–230.

[5] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan, "From taxi GPS traces to social and community dynamics: A survey," *ACM Comput. Surveys*, vol. 46, no. 2, 2013 Art. no. 17.

[6] T. Kindberg, M. Chalmers, and E. Paulos, "Guest editors' introduction: Urban computing," *IEEE Pervasive Comput.*, vol. 6, no. 3, pp. 18–20, Jul.-Sep. 2007.

[7] V. Kostakos and E. ONeill, "Cityware: Urban computing to bridge online and real-world social networks," *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. Hershey, PA, USA: IGI Global, 2008, pp. 195–204.

[8] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, 2011, pp. 89–98.

[9] S. Liu, W. Cui, Y. Wu, and M. Liu, "A survey on information visualization: Recent advances and challenges," *Visual Comput.*, vol. 30, no. 12, pp. 1373–1393, 2014.

[10] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, *Mastering the Information Age-Solving Problems with Visual Analytics*. Konstanz, Germany: Florian Mansmann, 2010.

[11] W. Chen, F. Guo, and F.-Y. Wang, "A survey of traffic data visualization," *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 6, pp. 2970–2984, Dec. 2015.

[12] Z. Wang, et al., "Visual exploration of sparse traffic trajectory data," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1813–1822, Dec. 2014.

[13] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1833–1842, Dec. 2014.

[14] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, "Visualizing interchange patterns in massive movement data," *Comput. Graph. Forum*, vol. 32, no. 3pt3, pp. 271–280, 2013.

[15] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Pölitz, "Discovering bits of place histories from people's activity traces," in *Proc. IEEE Symp Visual Analytics Sci. Technol*, 2010, pp. 59–66.

[16] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli, and R. Nair, "Allaboard: Visual exploration of cellphone mobility data to optimise public transport," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 2, pp. 1036–1050, Feb. 2015.

[17] W. Wu, et al., "TelCoVis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 935–944, Jan. 2016.

[18] S. Chen, et al., "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 270–279, Jan. 2016.

[19] A. M. MacEachren, et al., "SensePlace2: GeoTwitter analytics support for situational awareness," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2011, pp. 181–190.

[20] C. Vehlow, T. Reinhardt, and D. Weiskopf, "Visualizing fuzzy overlapping communities in networks," *IEEE Trans. Vis Comput Graph.*, vol. 19, no. 12, pp. 2486–2495, Dec. 2013.

[21] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. J. Zhu, and H. Qu, "Visual analysis of topic competition on social media," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2012–2021, Dec. 2013.

[22] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. M. Collins, "#FluxFlow: Visual analysis of anomalous information spreading on social media," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1773–1782, Dec. 2014.

[23] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti, "Interactive visual clustering of large collections of trajectories," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2009, pp. 3–10.

[24] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.

[25] FiveThirtyEight, "Is Uber making NYC rush-hour traffic worse?" 2015. [Online]. Available: http://fivethirtyeight.com/tag/uber/

[26] X. Huang, Y. Zhao, J. Yang, C. Zhang, C. Ma, and X. Ye, "TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 160–169, Jan. 2016.

[27] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni, "Visual analysis of route diversity," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2011, pp. 171–180.

[28] Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP), "NYC Taxi and Limousine Commission (TLC) Trip Record Data," 2015. [Online]. Available: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[29] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2565–2574, Dec. 2012.

[30] F. Wang, et al., "A visual reasoning approach for data-driven transport assessment on urban roads," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 103–112.

[31] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, 2011, pp. 316–324.

[32] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proc. 23rd ACM Int. Conf. Adv. Geographical Inf. Syst.*, 2015, Art. no. 33.

[33] C. Hurter, B. Tissoires, and S. Conversy, "FromDaDy: Spreading aircraft trajectories across views to support iterative queries," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1017–1024, Nov. 2009.

[34] R. Scheepens, C. Hurter, H. Van de Wetering, and J. J. van Wijk, "Visualization, selection, and analysis of traffic flows," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 379–388, Jan. 2016.

[35] K. Kloeckl, X. Chen, C. Sommer, C. Ratti, and A. Biderman, "Trains of data," 2011. [Online]. Available: http://senseable.mit.edu/trainsofdata/

[36] L. Chen, et al., "Container port performance measurement and comparison leveraging ship gps traces and maritime open data," *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 5 pp. 1227–1242, May 2015.

[37] R. Scheepens, N. Willems, H. Van de Wetering, G. Andrienko, N. Andrienko, and J. J. Van Wijk, "Composite density maps for multivariate trajectories," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2518–2527, Dec. 2011.

[38] H. Piringer, M. Buchetics, and R. Benedik, "ALVis: Situation awareness in the surveillance of road tunnels," in *Proc. IEEE Conf. Visual Analytics Sc. Technol.*, 2012, pp. 153–162.

[39] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, 2014, pp. 1027–1036.

[40] L. Yu, et al., "iVizTRANS: Interactive visual learning for home and work place detection from massive public transportation data," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2015, pp. 49–56.

[41] F. Calabrese and C. Ratti, "Real time Rome," *Netw. Commun. Studies*, vol. 20, no. 3/4, pp. 247–258, 2006.

[42] Sina Corporation, "Weibo APIs," 2015. [Online]. Available: http://open.weibo.com/wiki/API

[43] H. Bosch, et al., "ScatterBlogs2: Real-time monitoring of microblog messages through user-guided filtering," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2022–2031, Dec. 2013.

[44] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu, "Whisper: Tracing the spatiotemporal process of information diffusion in real time," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2649–2658, Dec. 2012.

[45] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin, "TargetVue: Visual analysis of anomalous user behaviors in online communication systems," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 280–289, Jan. 2016.

[46] J. Chae, et al., "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 143–152.

[47] M. Liu, S. Liu, X. Zhu, Q. Liao, F. Wei, and S. Pan, "An uncertainty-aware approach for exploratory microblog retrieval," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 250–259, Jan. 2016.

[48] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. J. Zhu, and R. Liang, "EvoRiver: Visual analysis of topic coopetition on social media," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1753–1762, Dec. 2014.

[49] Twitter, Inc., "REST APIs," 2016. [Online]. Available: https://dev.twitter.com/rest/public

[50] K. Wongsuphasawat and J. Lin, "Using visualizations to monitor changes and harvest insights from a global-scale logging infrastructure at twitter," in *Proc. IEEE Conf. Visual Analytics Sci Technol.*, 2014, pp. 113–122.

[51] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu, "OpinionFlow: Visual analysis of opinion diffusion on social media," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1763–1772, Dec. 2014.

[52] J. Zhao, L. Gou, F. Wang, and M. Zhou, "Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media," in *Proc. IEEE Conf. Visual Analytics Sci Technol.*, 2014, pp. 203–212.

[53] Y. Lu, et al., "Integrating predictive analytics and social media," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 193–202.

[54] N. Ferreira, et al., "Urbane: A 3D framework to support data driven decision making in urban development," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2015, pp. 97–104.

[55] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing New York City's noises with ubiquitous data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 715–725.

[56] J.-H. Haunert and L. Sering, "Drawing road networks with focus regions," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2555–2562, Dec. 2011.

[57] H. Doraiswamy, N. Ferreira, T. Damoulas, J. Freire, and C. T. Silva, "Using topological analysis to support event-guided exploration in urban data," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2634–2643, Dec. 2014.

[58] C. Palomo, Z. Guo, C. T. Silva, and J. Freire, "Visually exploring transportation schedules," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 170–179, Jan. 2016.

[59] Foursquare, "Foursquare API," 2014. [Online]. Available: https://developer.foursquare.com/

[60] R. Krueger, D. Thom, and T. Ertl, "Visual analysis of movement behavior using web data for context enrichment," in *Proc. IEEE Pacific Vis. Symp.*, 2014, pp. 193–200.

[61] H. Qu, W.-Y. Chan, A. Xu, K.-L. Chung, K.-H. Lau, and P. Guo, "Visual analysis of the air pollution problem in Hong Kong," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1408–1415, Nov./Dec. 2007.

[62] P. Accorsi, et al., "HydroQual: Visual analysis of river water quality," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 123–132.

[63] M. Sips, P. Köthur, A. Unger, H.-C. Hege, and D. Dransch, "A visual analytics approach to multiscale exploration of environmental time series," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2899–2907, Dec. 2012.

[64] W. Wu, Y. Zheng, H. Qu, W. Chen, E. Groller, and L. M. Ni, "Boundaryseer: Visual analysis of 2D boundary changes," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 143–152.

[65] S. Goodwin, et al., "Creative user-centered visualization design for energy analysts and modelers," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2516–2525, Dec. 2013.

[66] S. Goodwin, J. Dykes, A. Slingsby, and C. Turkay, "Visualizing multiple variables across scale and geography," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 599–608, Jan. 2016.

[67] T. Muhlbacher and H. Piringer, "A partition-based framework for building and validating regression models," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 1962–1971, Dec. 2013.

[68] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*. Berlin, Germany: Springer Science & Business Media, 2011.

[69] J. Li, K. Zhang, and Z.-P. Meng, "Vismate: Interactive visual analysis of station-based observation data on climate changes," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 133–142.

[70] Y. Tang, H. Qu, Y. Wu, and H. Zhou, "Natural textures for weather data visualization," in *Proc. 10th Int. Conf. Inf. Vis.*, 2006, pp. 741–750.

[71] M. Borkin, et al., "Evaluation of artery visualizations for heart disease diagnosis," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2479–2488, Dec. 2011.

[72] P. Klemm, et al., "3D regression heat map analysis of population study data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 81–90, Jan. 2016.

[73] M. H. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "TimeSpan: Using visualization to explore temporal multi-dimensional data of stroke patients," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 409–418, Jan. 2016.

[74] M. Termeer, J. O. Bescós, M. Breeuwer, A. Vilanova, F. Gerritsen, and M. E. Groller, "CoViCAD: Comprehensive visualization of coronary artery disease," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1632–1639, Nov./Dec. 2007.

[75] Z. Zhang, et al., "The five Ws for information visualization with application to healthcare informatics," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 11, pp. 1895–1910, Nov. 2013.

[76] J. Zhang, et al., "Visual analysis of public utility service problems in a metropolis," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1843–1852, Dec. 2014.

[77] MIT Senseable City Lab and Spanish bank BBVA, "Spring Spree: Spending patterns in Spain during Easter 2011," 2012. [Online]. Available: http://senseable.mit.edu/bbva/
[78] S. Sobolevsky, I. Sitko, R. Tachet des Combes, B. Hawelka, J. Murillo Arias, and C. Ratti, "Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. the case of residents and foreign visitors in Spain," in *Proc. IEEE Int. Congr. Big Data*, 2014, pp. 136–143.
[79] C. Xie, W. Chen, X. Huang, Y. Hu, S. Barlowe, and J. Yang, "VAET: A visual analytics approach for e-transactions time-series," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1743–1752, Dec. 2014.
[80] H. Ziegler, M. Jenny, T. Gruse, and D. Keim, "Visual market sector analysis for financial time series data," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 83–90.
[81] C. Shi, S. Fu, Q. Chen, and H. Qu, "VisMOOC: Visualizing video clickstream data from massive open online courses," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 277–278.
[82] D. Trimm, P. Rheingans, and M. DesJardins, "Visualizing student histories using clustering and composition," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2809–2818, Dec. 2012.
[83] J. Jo, J. Huh, J. Park, B. Kim, and J. Seo, "LiveGantt: Interactively visualizing a large manufacturing schedule," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2329–2338, Dec. 2014.
[84] C. Dietrich, D. Koop, H. Vo, and C. Silva, "Baseball4D: A tool for baseball game reconstruction and visualization," in *Proc. IEEE Trans. Visual Analytics Sci.*, Oct. 2014, pp. 23–32.
[85] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. Keim, and O. Deussen, "Feature-driven visual analytics of soccer data," in *Proc. IEEE Trans. Visual Analytics Sci.*, 2014, pp. 13–22.
[86] C. Perin, R. Vuillemot, and J.-D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2506–2515, Dec. 2013.
[87] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko, "SnapShot: Visualization to propel ice hockey analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2819–2828, Dec. 2012.
[88] T. Polk, J. Yang, Y. Hu, and Y. Zhao, "TenniVis: Visualization for tennis match analysis," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2339–2348, Dec. 2014.
[89] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*. Berlin, Germany: Springer Science & Business Media, 2013.
[90] Google, "Google Charts," 2016. [Online]. Available: https://developers.google.com/chart/
[91] Journalism++ Cologne, "Datawrapper," 2016. [Online]. Available: https://datawrapper.de/
[92] Baidu, "Baidu Echarts," 2016. [Online]. Available: http://echarts.baidu.com.cn/
[93] IBM, "Many Eyes," 2015. [Online]. Available: https://manyeyes.alphaworks.ibm.com/manyeyes/
[94] Tableau Software, "Tableau," 2016. [Online]. Available: http://www.tableau.com/
[95] J. Zhao, P. Forer, and A. S. Harvey, "Activities, ringmaps and geovisualization of large human movement fields," *Inf. Vis.*, vol. 7, no. 3/4, pp. 198–209, 2008.
[96] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic, "Time curves: Folding time to visualize patterns of temporal evolution in data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 559–568, Jan. 2016.
[97] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, "Reducing snapshots to points: A visual analytics approach to dynamic network exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 1–10, Jan. 2016.
[98] A. Cox and V. Nguyen, "Oils roller coaster ride," 2008. [Online]. Available: http://nyti.ms/yfZB8X
[99] S. Haroz, R. Kosara, and S. Franconeri, "The connected scatterplot for presenting paired time series," *IEEE Trans. Vis. Comput. Graph.*, 2015, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7332976&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7332976
[100] H. Rosling, "Gapminder," 2006. [Online]. Available: http://www.gapminder.org
[101] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1325–1332, Nov./Dec. 2008.
[102] M. Barry and B. Card, "Visualizing MBTA data," 2014. [Online]. Available: http://mbtaviz.github.io/
[103] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl, "TrajectoryLenses–a set-based filtering and exploration technique for long-term trajectory data," *Comput. Graph. Forum*, vol. 32, no. 3pt4, pp. 451–460, 2013.
[104] N. Andrienko, G. Andrienko, L. Barrett, M. Dostie, and P. Henzi, "Space transformation for understanding group movement," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2169–2178, Dec. 2013.
[105] S. Liu, J. Pu, Q. Luo, H. Qu, L. M. Ni, and R. Krishnan, "VAIT: A visual analytics system for metropolitan transportation," *IEEE Trans. Intell. Transport. Syst.*, vol. 14, no. 4, pp. 1586–1596, Dec. 2013.
[106] Y. Zheng, W. Wu, H. Qu, C. Ma, and L. M. Ni, "Visual analysis of bi-directional movement behavior," in *Proc. IEEE Int. Conf. Big Data*, 2015, pp. 581–590.
[107] G. L. Andrienko and N. V. Andrienko, "Interactive maps for visual data exploration," *Int. J. Geographical Inf. Sci.*, vol. 13, no. 4, pp. 355–374, 1999.
[108] T. Munzner, *Visualization Analysis and Design*. Boca Raton, FL, USA: CRC Press, 2014.
[109] K. Verbeek, K. Buchin, and B. Speckmann, "Flow map layout via spiral trees," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2536–2544, Dec. 2011.
[110] D. Guo and X. Zhu, "Origin-destination flow data smoothing and mapping," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2043–2052, Dec. 2014.
[111] M. Bostock, "Choropleth," 2012. [Online]. Available: http://bl.ocks.org/mbostock/4060606
[112] D. Chu, et al., "Visualizing hidden themes of taxi movement with semantic transformation," in *Proc. IEEE Pacific Vis. Symp.*, 2014, pp. 137–144.
[113] T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma, "Proximity-based visualization of movement trace data," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2009, pp. 11–18.
[114] M. Lu, Z. Wang, and X. Yuan, "TrajRank: Exploring travel behaviour on a route by trajectory ranking," in *Proc. IEEE Pacific Vis. Symp*, 2015, pp. 311–318.
[115] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li, "Geometry-based edge clustering for graph visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1277–1284, Nov./Dec. 2008.
[116] O. Ersoy, C. Hurter, F. V. Paulovich, G. Cantareiro, and A. Telea, "Skeleton-based edge bundling for graph visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2364–2373, Dec. 2011.
[117] D. Holten and J. J. Van Wijk, "Force-directed edge bundling for graph visualization," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 983–990, 2009.
[118] A. Lambert, R. Bourqui, and D. Auber, "Winding roads: Routing edges into bundles," in *Proc. 12th Eurographics/IEEE – VGTC Conf. Vis.*, 2010, pp. 853–862.
[119] R. Scheepens, N. Willems, H. van de Wetering, and J. J. Van Wijk, "Interactive visualization of multivariate trajectory data with density maps," in *Proc. IEEE Pacific Vis. Symp.*, 2011, pp. 147–154.
[120] H. Zhou, P. Xu, X. Yuan, and H. Qu, "Edge bundling in information visualization," *Tsinghua Sci. Technol.*, vol. 18, no. 2, pp. 145–156, 2013.
[121] J. Feinberg, "Wordle," 2008. [Online]. Available: http://www.wordle.net/
[122] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 1137–1144, Nov./Dec. 2009.
[123] D. Yang, D. Zhang, and B. Qu, "Participatory cultural mapping based on collective behavior data in location-based social networks," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, 2016, Art. no. 30.
[124] W. Cui, H. Qu, H. Zhou, W. Zhang, and S. Skiena, "Watch the story unfold with textwheel: Visualization of large-scale news streams," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 2, 2012, Art. no. 20.
[125] M.-J. Kraak, "The space-time cube revisited from a geovisualization perspective," in *Proc. 21st Int. Cartographic Conf.*, 2003, pp. 1988–1996.
[126] P. Gatalsky, N. Andrienko, and G. Andrienko, "Interactive analysis of event data using space-time cube," in *Proc. 8th Int. Conf. Inf. Vis.*, 2004, pp. 145–152.

[127] C. Tominski, P. Schulze-Wollgast, and H. Schumann, "3D information visualization for time dependent data on maps," in *Proc. 9th Int. Conf. Inf. Vis.*, 2005, pp. 175–181.

[128] U. D. Turdukulov, M.-J. Kraak, and C. A. Blok, "Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds," *Comput. Graph.*, vol. 31, no. 3, pp. 370–379, 2007.

[129] T. Kapler and W. Wright, "GeoTime information visualization," *Inf. Vis.*, vol. 4, no. 2, pp. 136–146, 2005.

[130] M. S. John, M. B. Cowen, H. S. Smallman, and H. M. Oonk, "The use of 2D and 3D displays for shape-understanding versus relative-position tasks," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 43, no. 1, pp. 79–98, 2001.

[131] D. Guo, J. Chen, A. M. MacEachren, and K. Liao, "A visualization system for space-time and multivariate patterns (VIS-STAMP)," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 6, pp. 1461–1474, Nov./Dec. 2006.

[132] N. Andrienko and G. Andrienko, "Interactive visual tools to explore spatio-temporal variation," in *Proc. Work. Conf. Adv. Visual Interfaces*, 2004, pp. 417–420.

[133] G. Sun, Y. Liu, W. Wu, R. Liang, and H. Qu, "Embedding temporal display into maps for occlusion-free visualization of spatio-temporal data," in *Proc. IEEE Pacific Vis. Symp.*, 2014, pp. 185–192.

[134] S. Liu, D. Maljovec, B. Wang, P. Bremer, and V. Pascucci, "Visualizing high-dimensional data: Advances in the past decade," in *Proc. Eurographics Conf. Vis.*, 2015, pp. 20151115.127–147.

[135] D. Keim, "Information visualization and visual data mining," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 1–8, Jan.-Mar. 2002.

[136] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne, "Flowstrates: An approach for visual exploration of temporal origin-destination data," *Comput. Graph. Forum*, vol. 30, no. 3, pp. 971–980, 2011.

[137] R. Krüger, D. Thom, and T. Ertl, "Visual analysis of movement behavior using web data for context enrichment," in *Proc. IEEE Pacific Vis. Symp.*, 2014, pp. 193–200.

[138] J. Poco, et al., "Visual reconciliation of alternative similarity spaces in climate modeling," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1923–1932, Dec. 2014.

[139] M. L. Pack, K. Wongsuphasawat, M. VanDaniker, and D. Filippova, "ICE–visual analytics for transportation incident datasets," in *Proc. IEEE Int. Conf. Inf. Reuse Integr.*, 2009, pp. 200–205.

[140] J. Lanir, P. Bak, and T. Kuflik, "Visualizing proximity-based spatiotemporal behavior of museum visitors using tangram diagrams," *Comput. Graph. Forum*, vol. 33, no. 3, pp. 261–270, 2014.

[141] Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang, and H. Qu, "egoSlider: Visual analysis of egocentric network evolution," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 260–269, Jan. 2016.

[142] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen, "Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2603–2612, Dec. 2012.

[143] M. O. Ward, "Multivariate data glyphs: Principles and practice," in *Handbook of Data Visualization*. Berlin, Germany: Springer, 2008, pp. 179–198.

[144] S. Feiner and C. Beshers, "Visualizing N-dimensional virtual worlds with n-vision," *ACM SIGGRAPH Comput. Graph.*, vol. 24, no. 2, pp. 37–38, 1990.

[145] J. LeBlanc, M. O. Ward, and N. Wittels, "Exploring N-dimensional databases," in *Proc. 1st Conf. Vis.*, 1990, pp. 230–237.

[146] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo, "TopicPanorama: A full picture of relevant topics," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2014, pp. 183–192.

[147] H. Qu, "Co-authorship graph," 2015. [Online]. Available: http://lbnx07.ust.hk/coauthor/coauthor?p=huamin

[148] J. Heer and D. Boyd, "Vizster: Visualizing online social networks," in *Proc. IEEE Symp. Inf. Vis.*, 2005, pp. 32–39.

[149] Q. Liu, Y. Hu, L. Shi, X. Mu, Y. Zheng, and J. Tang, "EgoNetCloud: Event-based egocentric dynamic network visualization," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2015, pp. 65–72.

[150] M. Balzer, O. Deussen, and C. Lewerentz, "Voronoi treemaps for the visualization of software metrics," in *Proc. ACM Symp. Softw. Vis.*, 2005, pp. 165–172.

[151] F. S. Duarte, F. Sikansi, F. M. Fatore, S. G. Fadel, and F. V. Paulovich, "Nmap: A novel neighborhood preservation space-filling algorithm," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2063–2071, Dec. 2014.

[152] D. A. Keim, "Solving problems with visual analytics: Challenges and applications," in *Proc. Euro. Conf. Mach. Learn. Knowl. Discovery Databases*, 2012, pp. 5–6.

[153] G. Andrienko and N. Andrienko, "Spatio-temporal aggregation for visual analysis of movements," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2008, pp. 51–58.

[154] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren, "MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 11–20, Jan. 2016.

[155] Z. Wang and X. Yuan, "Urban trajectory timeline visualization," in *Proc. Int. Conf. Big Data Smart Comput.*, 2014, pp. 13–18.

[156] K. Kloeckl, O. Senn, and C. Ratti, "Enabling the real-time city: LIVE Singapore!" *J. Urban Technol.*, vol. 19, no. 2, pp. 89–112, 2012.

[157] A. H. Meghdadi and P. Irani, "Interactive exploration of surveillance video through action shot summarization and trajectory visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2119–2128, Dec. 2013.

[158] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel, "From movement tracks through events to places: Extracting and characterizing significant places from mobility data," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2011, pp. 161–170.

[159] J. Poco, H. Doraiswamy, H. T. Vo, J. L. D. Comba, J. Freire, and C. T. Silva, "Exploring traffic dynamics in urban environments using vector-valued functions," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 161–170, 2015.

[160] R. Krueger, D. Thom, and T. Ertl, "Semantic enrichment of movement behavior with foursquare–a visual analytics approach," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 8, pp. 903–915, Aug. 2015.

[161] B. C. Kwon, S.-H. Kim, S. Lee, J. Choo, J. Huh, and J. S. Yi, "VisOHC: Designing visual analytics for online health communities," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 71–80, Jan. 2016.

[162] G. Grinstein, et al., "VAST 2011 challenge: Cyber security and epidemic," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2011, pp. 299–301.

[163] N. Nabian, C. Ratti, A. Biderman, and G. Grise, "MIT GEOblog: A platform for digital annotation of space for collective community based digital story telling," in *Proc. 3rd IEEE Int. Conf. Digital Ecosystems Technol.*, 2009, pp. 353–358.

[164] M. Szell, S. Grauwin, and C. Ratti, "Contraction of online response to major events," *PLoS ONE*, vol. 9, no. 2, 2014, Art. no. e89052.

[165] A. Dasgupta, J. Poco, Y. Wei, R. Cook, E. Bertini, and C. Silva, "Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 9, pp. 996–1014, Sep. 2015.

[166] J. Poco, et al., "SimilarityExplorer: A visual inter-comparison tool for multifaceted climate data," *Comput. Graph. Forum*, vol. 33, no. 3, pp. 341–350, 2014.

[167] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual Analytics: Definition, Process, and Challenges*. Berlin, Germany: Springer, 2008.

[168] J. Paiva, W. Schwartz, H. Pedrini, and R. Minghim, "An approach to supporting incremental visual data classification," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 1, pp. 4–17, Jan. 2015.

[169] B. Settles, *Active Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning)*. Long Island, NY, USA: Morgan & Clay Pool, 2012.

[170] G. Montavon, G. B. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012.

[171] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

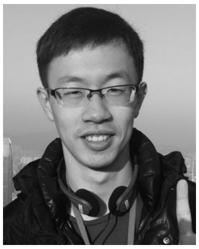
[172] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[173] S. M. Arietta, A. A. Efros, R. Ramamoorthi, and M. Agrawala, "City forensics: Using visual elements to predict non-visual city attributes," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2624–2633, Dec. 2014.

[174] P. Quinan and M. Meyer, "Visually comparing weather features in forecasts," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 389–398, Jan. 2016.
[175] Z. Liao, Y. Yu, and B. Chen, "Anomaly detection in GPS data based on visual analytics," in *Proc. IEEE Symp. Visual Analytics Sci. Technol.*, 2010, pp. 51–58.
[176] J. Wang and K. Mueller, "The visual causality analyst: An interactive interface for causal reasoning," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 230–239, Jan. 2016.
[177] N. Cao, D. Gotz, J. Sun, and H. Qu, "DICON: Interactive visual analysis of multidimensional clusters," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2581–2590, Dec. 2011.
[178] L. Lu, N. Cao, S. Liu, L. Ni, X. Yuan, and H. Qu, "Visual analysis of uncertainty in trajectories," in *Proc. 18th Pacific-Asia Conf. Adv. Knowl. Discovery Data Min.*, pp. 509–520, 2014.
[179] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE Trans. Big Data*, vol. 1, no. 1, pp. 16–34, Mar. 2015.
[180] J. Howe, "Crowdsourcing: A definition," *Crowdsourcing: Tracking the Rise of the Amateur*, 2006, http://crowdsourcing.typepad.com/cs/2006/06/
[181] B. Guo, C. Chen, D. Zhang, Z. Yu, and A. Chin, "Mobile crowd sensing and computing: When participatory sensing meets participatory social media," in *Proc. IEEE Commun. Mag.*, 2016, pp. 131–137.
[182] B. Guo, H. Chen, Z. Yu, X. Xie, S. Huangfu, and D. Zhang, "FlierMeet: A mobile crowdsensing system for cross-space public information reposting, tagging, and sharing," *IEEE Trans. Mobile Comput.*, vol. 14, no. 10, pp. 2020–2033, Oct. 2015.

Yixian Zheng received the bachelor's degree and master's degree from Shanghai Jiao Tong University, China, in 2010 and 2013, respectively. He is currently working toward the PhD degree in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests include visual analytics and visualization of multivariate data.

Wenchao Wu received the bachelor's degree from Shanghai Jiao Tong University in 2010, and the master's degree from School of Electrical and Computer Engineering, Georgia Institute of Technology in 2013. He is currently working toward the PhD degree in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests include visual analytics, data visualization, and human-computer interactions.



Yuanzhe Chen received the bachelor's degree and master's degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, in 2014 and 2011, respectively. He is currently working toward the PhD degree in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests include information visualization, and visual analytics and computer vision.

Huamin Qu is a professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests include visualization and computer graphics, with focuses on urban informatics, social network analysis, e-learning, and text visualization. His research has been recognized by many awards, including five best paper/honorable mention awards. He is a member of the IEEE.

Lionel M. Ni is chair professor in the Department of Computer and Information Science and vice rector for Academic Affairs at the University of Macau, and fellow of Hong Kong Academy of Engineering Science. He has chaired more than 30 professional conferences and has received eight awards for authoring outstanding papers. He serves on the editorial boards of the *Communications of the ACM*, the *IEEE Transactions on Big Data* and the *ACM Transactions on Sensor Networks*. He is a fellow member of the IEEE.