

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI-

Unfolding the City: Spatial Preference Based on Individual Demographic Characteristics

QI WANG^{1,2}, MIN LU^{2,3} and QINGQUAN LI^{1,2}

¹State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

²Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

³School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China

Corresponding author: Qingquan Li (e-mail: liqq@szu.edu.cn).

This work was jointly supported by the National Natural Science Foundation of China(No. 71961137003 and No. 41671387).

ABSTRACT A city is shaped not only by its assembled infrastructures but also the people living inside it. People in a wide range of individual demographic characteristics visit places with different weights and preferences, sculpturing the city into a space of diverse spatial preferences. Different from the well study of anonymous travel behaviors, the daily movement of characterized individuals, is less studied due to the challenges in collection and analysis of data with privacy. In this work, we take a step forward and perform an online census to collect individual profiles and movements from large-scale volunteers. Then we design a visual analytic system to investigate the spatial preference to groups by specifying demographic characteristics. To facilitate the identification of characterized groups, individuals are embedded in t-SNE projection for an abstract overview as well as drawn as vivid graphics by a data-driven profile method for detail examination. A 2.5D spatial visualization is proposed to maintain a compact multivariable analysis by relaxing the z-axis to encode information, such as visiting frequency, demands, traveling distance, etc. Together with the cross-filter and flexible 2.5D interactions, the effectiveness and usability of the system are well demonstrated by a study made in Shenzhen.

INDEX TERMS Visual analytics, spatial preference, individual characteristics, data collection

I. INTRODUCTION

THE science fiction **Folding Beijing** [1] depicts a world where three social classes of people live in independent spatiotemporal patterns, though sharing the same earth surface. This is an artistic metaphor for one potential fact that people prefer to select different places to live according to individual characteristics, especially personal demographic characteristics. This phenomenon has attracted the research interest for a long time in the sociological field. Back to 1970, Pahl [2] posed the question "whose city" and interest in understanding the territorial inequalities. Contemporary contestation also continually raises the banner "The City Belongs to All!" [3]. Some approaches have been to discuss spatial distribution [4], spatial interaction [5] and place attractiveness [6]. However, lacking of movers' characteristics, they almost only analyzed from the spatial perspective, and individual characteristic factors were not put into the context of the problems. It is required to explore the relationship between spatial preference and individual demographic char-

acteristics for better region utilization, diversity mix-up, and opportunity promotion.

Due to the multidimensional nature of demographic characteristics and spatial diversity of movements, it is challenging to explore, analysis and express the relationship between them. In the past decades, the power of visual analytics on urban problems has been recognized and well developed [7], [8]. The presence of social media applications fuels the spatial analysis with rich semantic texts, so researchers get the chances to explore the thematic meaning of the moving behavior, such as events [9] or topics [10]. Those works essentially contribute on deepening the contextual understanding of moving activity.

However, they rarely combine the movement analysis with individual characteristics for better region utilization, diversity mix-up, and opportunity promotion. The question of how to intuitively discover the relationship between spatial preference and multidimensional characteristics is worthy of further investigation.

In this work, we contextualize the analysis of citizens' spatial preference in the backcloth of individual demographic characteristics for better region utilization, diversity mix-up, and opportunity promotion.

characteristics. With the motivation of it, we need to obtain the demographic characteristics of individuals and their daily movements. The conventional offline census takes lots of effort and too much time to be out of date. With the wide use of the mobile device, it is possible to collect information online. In this work, we ran an online demographical collection campaign released via a social media application. Based on the data collected, we developed a visual analytics system to explore where do specified characterized citizens prefer to go. The system incorporates visualizations of individuals' characteristics, including t-SNE for an overview and data-driven profile visualization for the detailed check, as well as 2.5D spatial visualization to explore the spatial distribution of movements. Finally, applying our method in Shenzhen, several cases are derived to demonstrate the power of our method. There are two contributions of this work:

(1) investigate spatial preference with individual demographic characteristics, which is a natural step into research of social understanding of the city.

(2) develop a visual analytic system, whose effectiveness and usability is well demonstrated by a study made in Shenzhen.

II. RELATED WORK

This work concerns the research topic of spatial data visual analytics. Lots of related work focus on the spatiotemporal analysis in movement data, in the absence of thematic information. The presence of geotagged social media data flourishes the exploration of thematic information along with movement. By analyzing the texts, those work explain what drive the moving activities or what is resulted from. This work continues the thematic research in movement data but focuses more on individual demographic characteristics. Landed by a census experiment, this work is able to analyze the profile information directly, to get rid of indirect data inferred from social media data as other related works do.

A. MOVEMENT DATA ANALYSIS

With the development of location-acquisition techniques, massive spatial trajectories are collected, to keep track of the trajectories of various moving objects. Many techniques have been proposed to process, mine trajectory [11]. In the field of visualization and visual analytics, spatial visualizations are specifically designed for the time, locations, spatial-temporal information and other properties in the traffic data [12]. A large number of visual analytics tools and applications cover situation-aware exploration, pattern discovery and traffic situation monitoring. Wang et al [7] extract the traffic jam propagation graph extraction to reveal underlying data patterns. Guo et al. [13] and Zeng et al. [8] construct geographical regions and visually aggregate the in-between movements as flows. To discover the route travel patterns, Lu et al. [14] propose TrajRank to explore the route travel behavior based on ranking. For multiple routes, Liu et al. [15] study the route diversity between locations and Lu et al. [16] explore the route choice behavior among multiple routes.

B. GEO-TAGGED SOCIAL MEDIA DATA ANALYSIS

As the presence of social media services, social media data with geo-tags are collected to track people's movements in daily lives. As an analog of remote sensing data in social science research, the geospatial big data has been proposed as social sensing [17]. Hence, analyzing movement information along with rich text has become a hot research area in recent years. Those works infer thematic information from the semantic texts. Cao et al. [18] propose *Whisper* for tracing the pathways of tweets on a spatial hierarchical layout, to investigate how information flow among multiple places. Krueger et al. [19] used GPS and location-based service data to support the analysis of movement behaviors. Chen et al. [20] present a visual analytics system to support the exploration of sparsely sampled trajectory from social media. Some other researchers tend to infer real information, such as names, gender, etc, from social media data [21], to break down the demographic characteristics of social media users. Luo et al. [22] derive race, gender, and age as three demographics dimensions to analyze its impact on the urban human mobility patterns. Since real information is not enforced in social media, the wrong inferring is inevitable. Longley et al. [23] [24] identifies and assess the biases inherent in social media usage in social research and evaluate the deployment of social media data in research applications.

III. MOTIVATION

This work was sparked from a meeting with two domain experts who work on human mobility and urban informatics in the geospatial field. The experts pointed out that as big and open data evolved in, the scope of urban research has been lifted from the spatial perspective up to social-spatial perspective. There is a rising research interest in understanding how categories of citizens with certain demographic characteristics visit places during their daily lives. Whether people select places with preference which are intrinsically affected by their social-economic characteristics? Is there a limitation or probabilities of interaction between different socio-economic groups in public space? With the advances of data acquisition and analysis techniques, they demanded to go beyond the conventional static sampled questionnaires, which usually survey visiting preferences. They sought for the possibility in revealing the spatial distribution in real-world visiting logs and urged to have a dynamic system which supports the interactive exploration of spatial preference in different contexts of demographic characteristics.

Thereafter we booted up this project with several rounds of discussions, to frame the research problem with two core contents and the research procedure. The first content is **demographic characteristics**. Demographic characteristics of a person or a group of individuals are the descriptors which contribute to the specification of the populations. This work intends to cover the characteristics from multi-facets, including descriptors such as gender, income, education, etc. The second content is **spatial preference**. It is reflected by the popularity of a place to be visited.

Motivated by their research interest in understanding the spatial preference with individual demographic characteristics, this work goes through a procedure consists of two stages, i.e., data collection and basic reviewing, visual analytics design iteration with experts' feedback.

IV. DATA DESCRIPTION

Considering there is no existing public dataset could fit our topic well, this work is consolidated with a self-maintained data sourcing. We release an online information survey to collect the demographic characteristics and movement logs of individuals for almost half a year. Then a series of statistical analysis is performed on the collected data for sampling bias checking.

A. ONLINE SURVEY

The advent of mobile sensing techniques and social media applications makes it possible to collect spatial data by apps within the phones. Complementary to the conventional survey, it brings the benefit of convenience without much manpower and precise spatial locations by GPS instead of semantic description.

In this work, we perform the survey in Shenzhen, which is one of the most modern metropolis in China. The experiment is deployed on Wechat, the most widely used social media mobile application in China. Fig. 1 shows the interfaces of survey app. For privacy issue, all detailed personal information are desensitized to categorical levels. Considering the main aim on the relationship between spatial preference and demographic characteristics, two crucial aspects need to be collected:

Demographic Characteristics Fig. 2 lists the eight domains to give a generalized description of the individuals. The profile will serve as the ingredients for the analysis of mobility patterns over diverse individuals.

Traveling Trips Individuals can upload dynamic traveling trips (Fig. 1(b)). Each trip requires the information of start/end location, start/end time, traveling purpose, and traveling mode. Locations are recorded by GPS modules in the phones. To encourage the trip uploading, a credit system retains the contribution of individuals on trips and rewards the volunteers with the top credits (Fig. 1(c)).

B. STATISTICS OF COLLECTED DATA

Over the releasing time period from 2015-11 to 2016-01, 21,435 individuals participated in the census and 229,155 trips were collected.

Our case-study data is confined to a small proportion of Wechat users who opt to contribute their demographic profiles and movements. Considering the caveat that self-selecting individuals are most unlikely to represent any clearly defined population [23], we performed a series of preliminary statistics. In Fig.??, population distribution of each characteristic is shown. The length of the block is coded for population quantity.



FIGURE 1: Census app interface: (a) personal characteristics collecting page; (b) trips collecting page; (c) credit system page. (In real census survey, contents are shown in Chinese.)

1. Gender	6. Income
1-1 Female	6-1 <= 100k: whose annual pay is under 100k
1-2 Male	6-2 100-200k: whose annual pay is from 100k to 200k
2. Age	6-3 200-300k: whose annual pay is from 200k to 300k
2-1 <= 18: whose age is under 18	6-4 300-500k: whose annual pay is from 300k to 500k
2-2 19-40: whose age from 19 to 40	6-5 > 500k: whose annual pay is above 500k
2-3 41-60: whose age from 41 to 60	
2-4 > 60: whose age is over 60	
3. Residence	7. Car
3-1 Yes: who has registered permanent residence	7-1 Yes: who at least has one car
3-2 No: who doesn't have registered permanent residence	7-2 No: who has no car
4. Education	8. Job
4-1 Middle School: who highest education is or under middle school	8-1 Manager: who is responsible for administering in a company
4-2 High School: who highest education is high school	8-2 Technician: who does practical work in a laboratory or company
4-3 Undergraduate: who receives a bachelor degree	8-3 Businessman: who works in business or commerce
4-4 Graduate: who receives a Master or Ph.D degree	8-4 Servant: who performs duties for others
5. Estate	8-5 Workman: who works in labour
5-1 Yes: who has real estate	8-6 Government Officer: who works for government
5-2 No: who doesn't have real estate	8-7 Student: who is learner in school
	8-8 Retiree: who retired
	8-9 Other

FIGURE 2: Profile of Individual: eight individual characteristics enrich the analysis of mobility patterns.

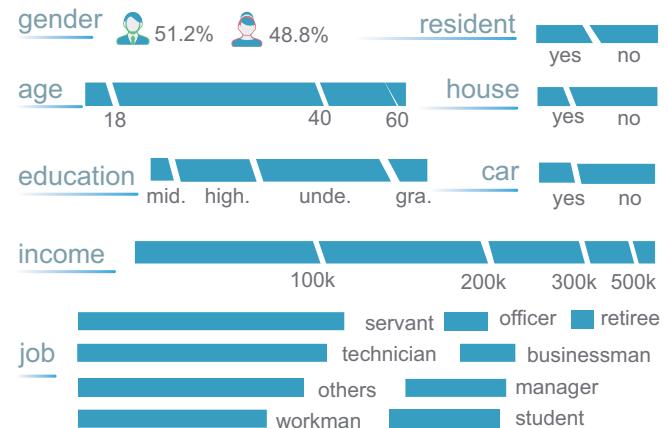


FIGURE 3: Preliminary statistics of each characteristic in collected demographic data.

From the figure, 51.2% of the participated citizens are male and 48.8% are female. It is consistent with governmental statistics that 47% of the whole citizens are female and 53%

are male¹. The age collected covers a wide range, dominating between 18 and 45. There are also records pertaining to individuals below the age of 18 or above 70. According to the 2015 Annual Census Statistics report², people aging 15-64 occupy 83.23% and the median age is 31.5. The age distribution in collected data follows the fact that Shenzhen is a city where the majority is young. From the distribution of education levels, ranging from low to high. The technical college and university dominate the samples at the 61% occupancy rate. The majority get paid below 200,000. Citizens with higher salary are also reached in our census. Due to lacking of governmental statistics, we cannot verify all characteristics collected are consistent with all the citizens in Shenzhen. From the perspective of survey, there is always inevitable bias inherent in fully representing the ground truth of the population. However, the preliminary statistical analysis shows the rich scope of the data, which embodies positive sign of a relatively even sampling of the population in the city.

The basic statistics related to movements are shown in Fig. 5. Fig. 5(a) visualizes the distribution of trip quantity of each citizen. It has a heavy-tail trait, which follows the law of most human-related statistics. In Fig. 5(b), the active traveling time (here, we take the start time as the representative of active time) follows the common knowledge about urban life. There are obvious morning and evening rush hours. Fig. 5(c) gives the counting of different traveling purposes. 95% trips are tagged with clear traveling purpose in the data. 33% are going home and 37% are going to work. Besides this kind of routine traveling, there are also substantial trips such as going shopping, going the hospital, etc. Fig. 5(d) shows the spatial distribution of the origins and destinations. It is found that more dots are located in Futian and Nanshan districts, the city's heart than the surrounding areas. This is consistent to Batty's exposition of the focus of city networks and interaction patterns [25].

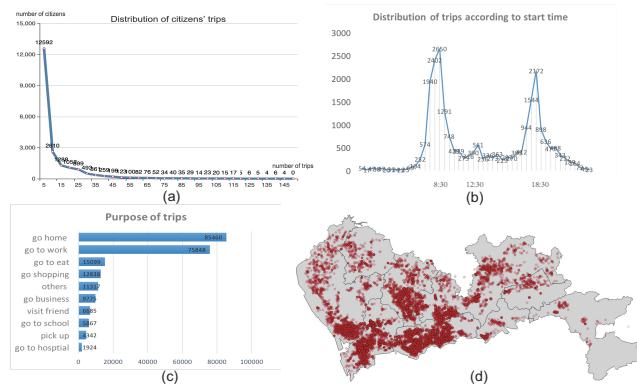


FIGURE 4: Statistics of Trips: (a) distribution of trip quantity of each citizen; (b) start time of trips; (c) purpose of trips; (d) distribution of origins/destinations.

¹http://tjj.sz.gov.cn/zwgk/zfxxgkml/tjsj/tjgb/201804/t20180427_11800902.htm
²http://www.sztj.gov.cn/xxgk/tjsj/pccb/201606/t20160614_3697000.htm

V. VISUAL DESIGN

The design of visual analytics system is iterated via multi-pass discussions with the two domain experts. In this section, we first introduce specific tasks the system intended for and design rationales aligned to provide the user-friendly interface for domain experts. Then we introduce the visual design.

A. TASKS

The exploration of core contents (introduced in Section III) is resolved further with the following tasks.

- **Task 1: Identify the group with specific individual demographic characteristics:** to get groups of people with common or similar attributes, to explore the correlation among individual characteristics.
- **Task 2: Explore the spatial preference of a group:** to explore where, how and why an interested group go in the city.
- **Task 3: Compare spatial preference among multiple groups:** to express the similarities and differences between different groups in their spatial preferences, to investigate the relationship between movement and individual demographic characteristics.

B. DESIGN CONSIDERATIONS

With three tasks above, we derive following design considerations:

- **Intuitive perception of an individual as an organic complex (C1):** the system should make use of end-users' daily experience in sensing individuals' demographic characteristics to provide intuitive visualization, instead of the lifeless representation by number. The visual design needs to help end-users to pick desirable ones from the mass.
- **Good overview of the multivariate individuals (C2):** following the visual analytics manta by Shneiderman [26], it is very important to provide a good overview of all the individuals then the users know where to explore.
- **Effective multivariable cross-filter for individual characteristics (C3):** there are eight domains to describe an individual. The system is supposed to provide a straight-forward way for easy filtering by the eight criteria.
- **Compact visualization of mobility patterns in the constraint of spatial space (C4):** the analysis of mobility patterns not only includes the conventional spatial and temporal dimensions, but also other abstract dimensions, e.g., travel purpose, visiting frequency, etc. The system should handle a compact layout to support the easy correlation between spatial and abstract information.
- **Flexible interactions to explore the mobility patterns either within one group or between groups (C5):** to support the comparison among groups, **Task 3**, the sys-

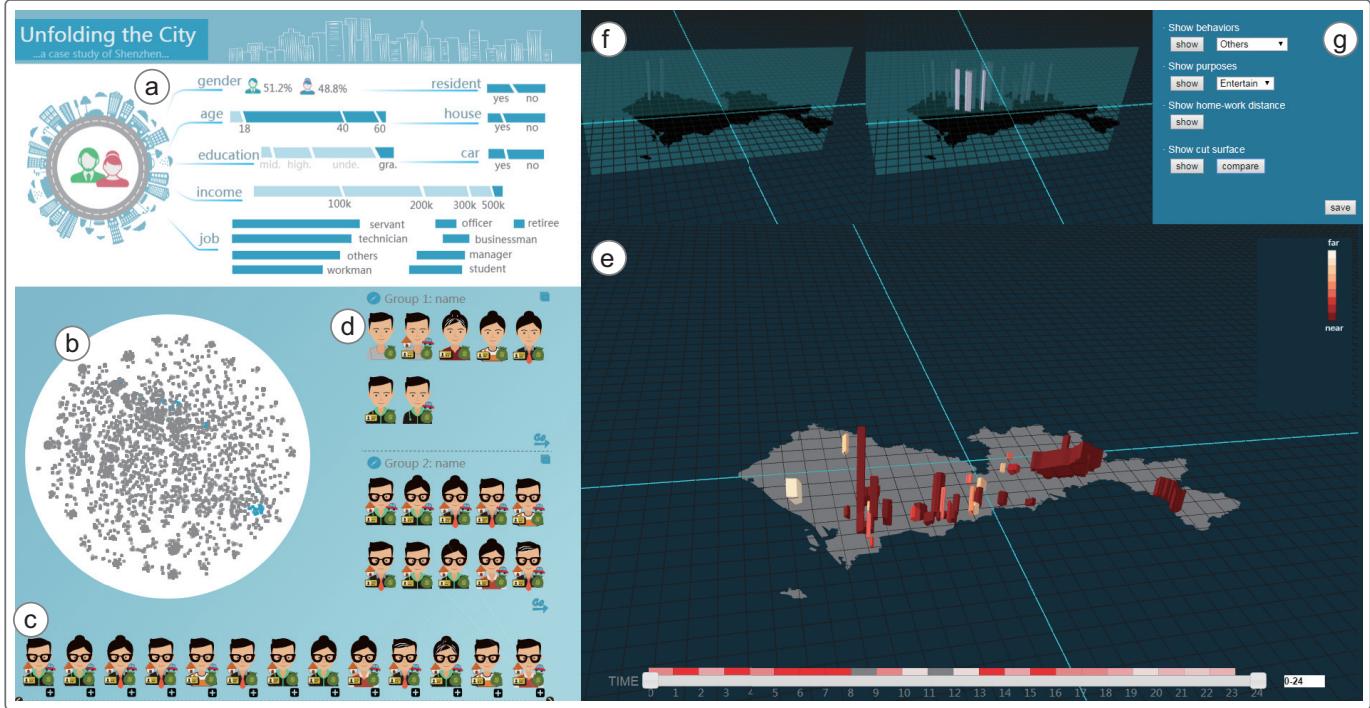


FIGURE 5: System Interface: (a) an interactive infographics provides the basic statistic facets of individuals; (b) a t-SNE visualization gives an overview of individuals, where the similarity in high dimensional space is preserved in 2D space; (c) data-driven profiles support users with organic individual perception; (d) group of interested individuals can be saved; (e) movement of the chosen group is visualized and explored in 2.5D main view; (f) a dock is provided to restore and compare findings; (g) different choices of 2.5D encoding can be selected.

tem should maintain flexible interactions which allow the end-users to explore freely.

C. VISUALIZATIONS

With all the design considerations, we develop a visual analytics system as Fig. 6 shows. It composes of individual panel (the left part) for individuals (*Task 1*) and spatial panel (the right part) for the mobility pattern (*Task 2 and 3*). In the individual panel, interactive infographics, t-SNE visualization, and data-driven profiles support users to narrow the scope down to groups of individuals with interested characteristics. With the specified group of individuals, its moving-related information is visualized and explored in a 2.5D spatiotemporal view.

1) Data-driven Profile Visualization

Glyph-based visualization [27] is the form of visual design to compose multivariable into a collection of unified visual symbols, known as a glyph. A glyph is intended for quick understanding and aligned comparison. Among glyph design, Chernoff Face [28] represents data variables by the different features of a cartoon face. Following the idea of Chernoff Face, we design a type of glyph, a graphical representation of people with specific individual characteristics. The idea behind using faces is that humans easily recognize faces and notice small changes without difficulty (*CI*). Those visual

profiles are intended for intuitive visual understanding, from abstract to concrete and semantic understanding, to support users to target the interested individual groups effectively.

Fig. 7 shows the legend for the user profile. The eight domains are encoded by visual symbols and organically organized in a human figure. There are basically three types of variables to drive the figure, i.e., the numeric, categorical and boolean ones. For numeric attributes such as income and education levels, visual symbol keeps consistent design but with changing visual variation, such as size, thickness. For the categorical attributes such as job, visual symbols are designed separately for better semantic meaning. For the boolean attribute like car, house, a symbol is designed to indicate its existence. With this consideration, the domains are designed as follows:

- **Gender** It is visually mapped to the hairstyle of the avatar.
- **Age** Age is implied by the decoration on the hair. For the elder above 70, the hair is dyed to grey. For the youth beneath 18, a hair decoration is adapted for the different hairstyles of girls and boys.
- **Education** The thickness of eyeglasses is used to indicate the different levels of education.
- **Job** The clothes is designed to imply the job of the individual. There are 9 types of clothes.
- **Belongings** For real estate, car, residential license are

considered as the belongings to the individual, so we design each of them as an add-on decoration to imply whether the individual has it or not.

- **Income** A money symbol is used to represent the income, whose size encodes the income level. The more money an individual earns, the larger the symbol is.

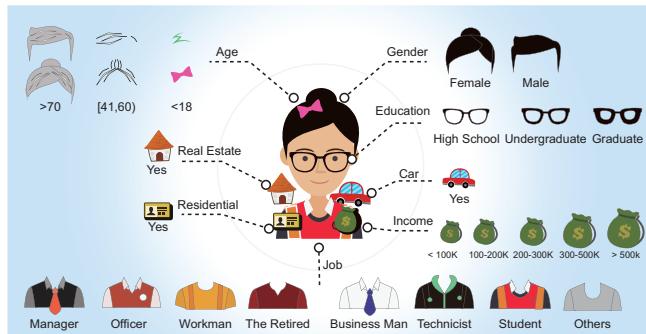


FIGURE 6: Design Profile: the eight characteristic domains are encoded and organized in an organic human figure.

With the visual mapping, the profile varies from individual to individual. By concretizing the attributes which otherwise is too abstract to percept, users can scan and search for interesting target effectively. Fig. 8 lists the figures with the top 10 largest population in some job. The textual number beneath indicates the population. It is found that the majority of workmen earn a low salary and most of them have no residential license. On the contrary, for the officers, all of them have the residential license and most of them have a house. Some of the retired people are in old age. Most of the managers are at undergraduates, even graduates.

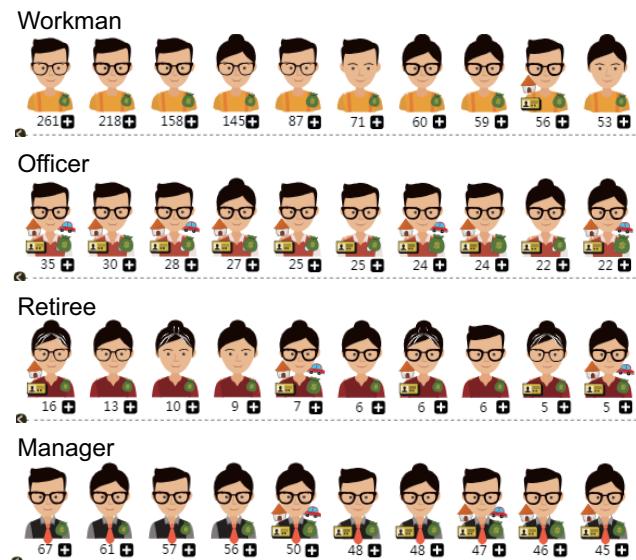


FIGURE 7: Representative figures with the top 10 largest population in four jobs

2) t-SNE Projection of Individuals

In our case, each individual can be denoted as a vector x_i with eight factors, and we get high dimensional data set $X = x_1, x_2, \dots, x_n$. The dimensionality reduction techniques are to preserve the local structure of high dimensional data in low dimensional space, which are efficient approaches to provide a good overview of the multivariable individuals (C2). We adapt t-SNE project [29] to project X as two-dimensional dots $Y = y_1, y_2, \dots, y_n$. One of steps in t-SNE is to compute the conditional probabilities to represent the similarity based on the distance between high dimensional datapoints. The conditional probability $p_{j|i}$ between x_j and x_i is given by:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\alpha_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\alpha_i^2)} \quad (1)$$

where α_i is the variance of the Gaussian that is centered on x_i . Specifically in our context, the high-dimensional Euclidean distances $\|x_i - x_j\|$ between x_i and x_j needs to be adapted the numerical and categorical characteristics. Characteristics such as age, income, are numerical and comparable, so the difference exactly explains when they are different. But the other characteristics, i.e. job, real estate, car, residential, are ordinal. There is not the numeric order. For example, the job distance from a manager to a businessman and a workman is not comparable, which is considered the same distance, i.e., set to 1.

As Fig. 9 shows, all 21435 volunteers are embedded in the 2D view, where the closer two dots are, the more similar they are in the eight characteristics. Fig. 9 exemplifies four features groups of dots in a neighbor.

Multiple views of abstract view, t-SNE protection, and semantic data-driven profile visualization are coordinated in a Cross-filter mechanism [30]. It allows end-users to interactive drill-down into individuals with interested characteristics from multiple perspectives(C3). Starting from the abstract criterion constraints, the scope of interest is narrowed down to individuals with(out) certain characteristics. And then further cross-filtering with semantically visual profiles can be performed to check the combination of 8 characteristic variables.

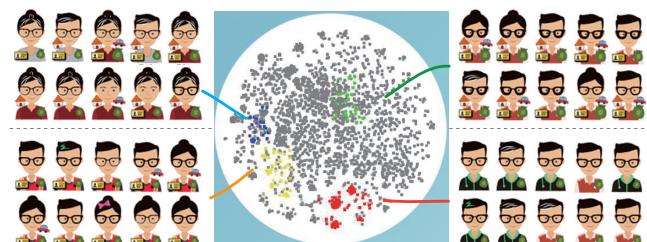


FIGURE 8: t-SNE project with four groups of the interest: (a) t-SNE project

3) 2.5D Spatial Visualization

Embedding multiple variables in the spatial map is a challenging problem. Distorting 2D spatial for better space usage,

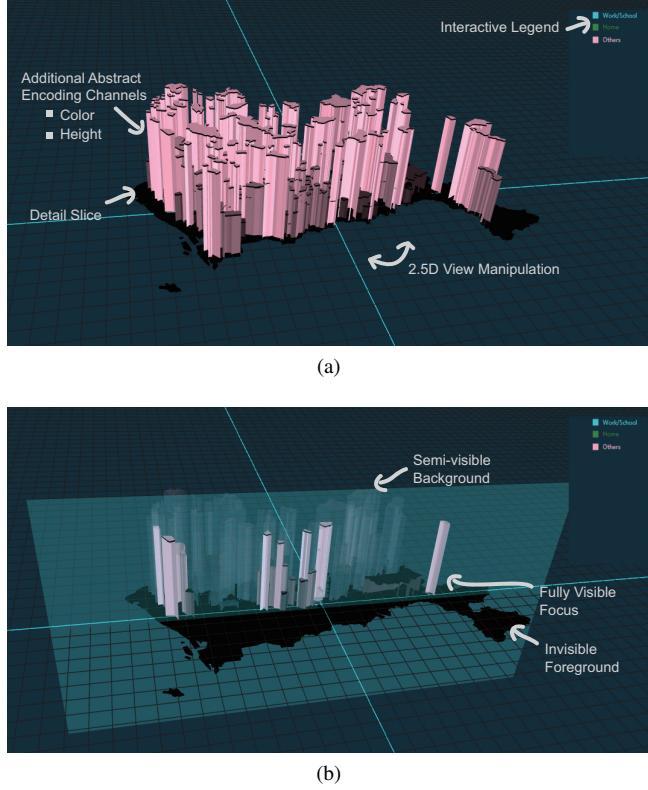


FIGURE 9: 2.5D Spatial Visualization: (a) 2.5D spatial visualization with z-axis to encode addition information; (b) detail slicing in 2.5D space.

such as the partial route embedding [31] is an effective method when there are a few local foci. When it comes to the global visualization, there is always a trade-off between occlusion-free and the compact spatial perception. Following the idea of stacking 2.5D space design [32], the space in this work is embedded in 2.5D space, to relax the z-axis for additional abstract information encoding.

Fig. 10(a) shows the 2.5D visualization. The spatial map consists of Traffic Analysis Zone (TAZ), which is the unit of geography holding a certain number of people. For the 2.5D, each TAZ is grown as a prism whose height and color are developed to encode abstract information, e.g., the occurrence of visiting. With difference encoding choices, it is capable to visualize the correlation between spatial and different attributes (*C4*). For example, in Fig. 10(a), prisms' height and color both encode the ratio of none-routine trips (such as go entertainment, go hospital, etc.) in the whole. For better 3D visual perception, lighting is turned on. Also, a lid is laid on the top of each prism to clearly bound its extent at the z-axis.

Several interactions are integrated into the spatial view to alleviating the occlusion problem (*C5*). The manipulation of 2.5D view supports users with rotate, zoom, pan operations, which is continuously flexible that users can find a suitable observing point. In the cases where occlusion is inevitable, a 2D view for overview and

an interaction, namely, Detail Slicing, are developed, as Fig. 10(b) and (c) shows. Overview view shows all TAZs in the foreground. It is for quickly sensing the distribution of the whole city. Detail Slicing is the operation to put a cross-section in the 2.5D space to expose the TAZ inside. To contextualize the cross-section, TAZs are visually adapted according to the intersection by ray casting. The TAZ in the foreground will be hidden, the intersected TAZs will be fully visualized and the ones in the background will be in semi-transparency.

To compare the mobility patterns across different groups of people, a Small Multiple dock is used to reserve the ever explored interesting result (Fig. 6(f)). To simply the comparison over spatial distribution, the snapshot is also rendered in the 2.5D space, which maintains the same interactions as the main 2.5D space does.

VI. CASE STUDIES

In this section, we apply the method described above to study the case of Shenzhen. We introduce the several cases to demonstrate the usage and effectiveness of the system.

A. CASE 1: HOME-LIFE DISTRIBUTION VS INCOME/EDUCATION

This work is inspired and motivated by the implication in science fiction *Folding Beijing*, which is that people are categorized into three classes and behave in the parallel spatial-temporal patterns. The first cast explores whether the folding city exists in reality or not.

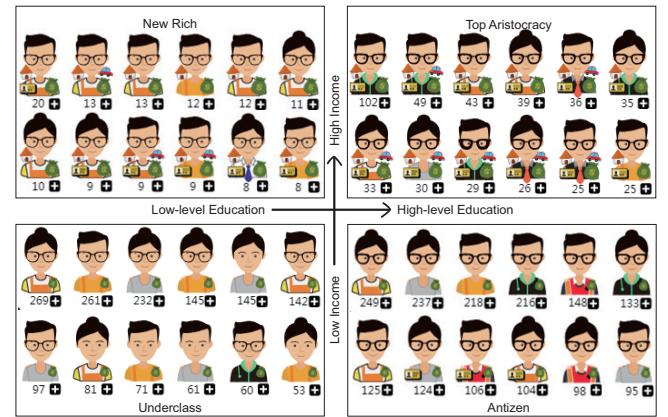


FIGURE 10: Four Groups Divided by Income and Education Characteristics

Although it is incomplete, income and education are two of the most essential characteristics to tag an individual with a good resource or not. As Fig. 11 shows, taking income and education as the two dimensions, taking 300,000 yuan as the border between high and low income and an undergraduate bachelor degree as the border between high and low education level, four different quadrants are generated in the four quadrants. In the first quadrant, individuals are rich and well educated, who are in a group of so-called **top aristocracy**. The second quadrant holds the group of **new**

rich, i.e., individuals with high income but low education income. In the third quadrant, they are the **underclass** who has both low income and education level. The fourth quadrant is **antizen**, which is a new word to describe people have a good education but earn little.

Fig. 11 shows the top 12 profiles with the maximum population in each group. In each group, other characteristics surprisingly demonstrate the high correlation to income and education. For example, almost all of the **top aristocracy** is with the residential license, house, and car. On the contrary, **underclass** do not have the residential license residence, neither the house nor the car. Most of **new rich** has a house and work in the service industry.

Before diving into the complex mobility pattern analysis, we explore the home distribution of the four groups. As Fig. 12 shows, **top aristocracy** and most of **new rich** live concentrated in the downtown area of Shenzhen, where the living cost is higher, especially the housing cost. A small part of **new rich** lives far away from the downtown, maybe because they don't have to work in downtown. The **underclass** distribute more dispersively all over the whole city. Compared to the rich, they prefer the outer space because of the lower living cost. Lots of **antizens** live in Nanshan District, where there are many universities and high-tech industrial parks full of well educated people.

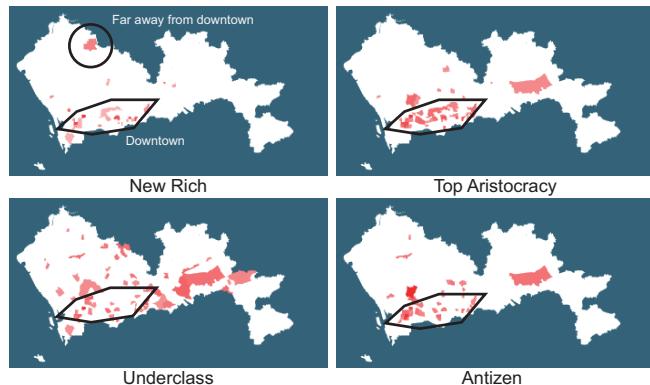


FIGURE 11: Home Distributions of the Four Groups

To have a better knowledge of mobility patterns of the four groups, trips are categorized into three categories, “home”, “work” for regular trips with the purpose of going home, work (or school for students), “other” for irregular trips with the purpose of shopping, visiting friends. Fig. 13(a) shows the 2.5D overview of the three kinds of traveling categories. Each TAZ is grown to the same height, i.e., the unit one, and the height proportion of each traveling categories encodes the percentage of certain purpose in the whole, pink for “others”, blue for “work” and green for “home”. Although inside TAZ prisms are blocked in the Fig. 13, our attention is drawn to the lifestyle of **New Rich**. It is founded that the **New Rich** travel for “other” purpose a lot, compared to other groups. To explore further, the detail slicing (introduced in Section V-C3) is applied to the four groups. Fig. 13(b)

shows three slices. It can be seen that individuals with low income (**underclass** and **antizen**) spend relatively more time on work, especially the **underclass**. Working percentage is very small in-group **new rich**. They live a more diverse lifestyle because they do many other things (the pink ones) except working.

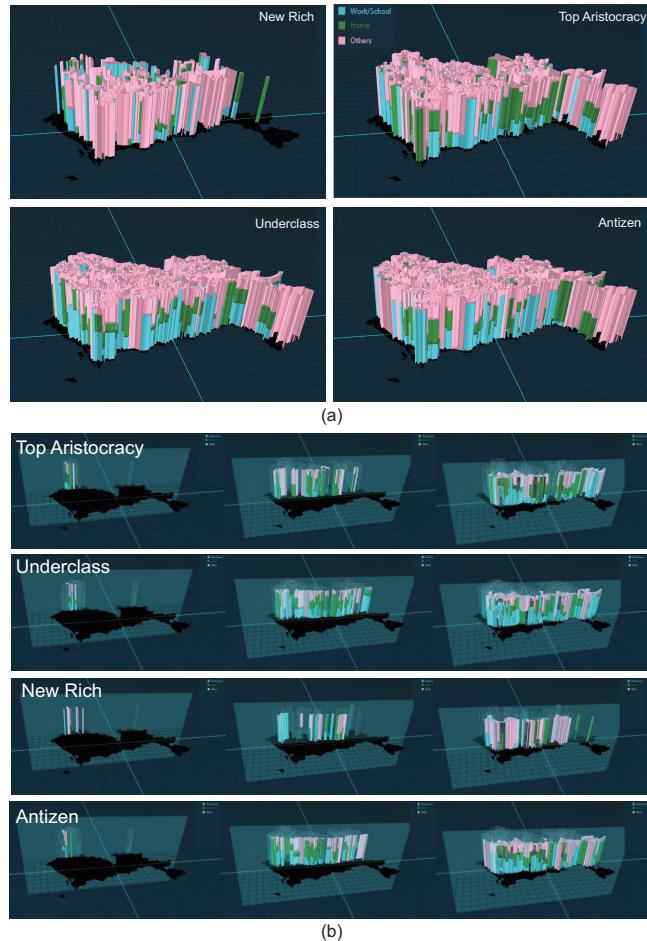


FIGURE 12: Mobility Patterns of “work”, “home” and “not essential” of the Four Groups: (a) the overview of the three slices; (b) three slices for detailed comparison.

B. CASE 2: SPATIAL SELECTION ON WORK-HOME VS INCOME/HOUSE

Home and work places are two crucial places where people stay long during the daily lives. Within it, home-work distance is an essential metric in home/work places selection. Meanwhile, it is influenced by many factors like transportation, housing, land use, etc. Understanding the home-work distance of a region helps to evaluate whether the region is well planned or not. In this case, we demonstrate our system's help in home-work distance understanding.

Considering income potentially determines where people live, we use income as the characteristics to select three groups, the group over 500K, the group during 200-300K and the group less than 100K. In Fig. 14, the height of the bar

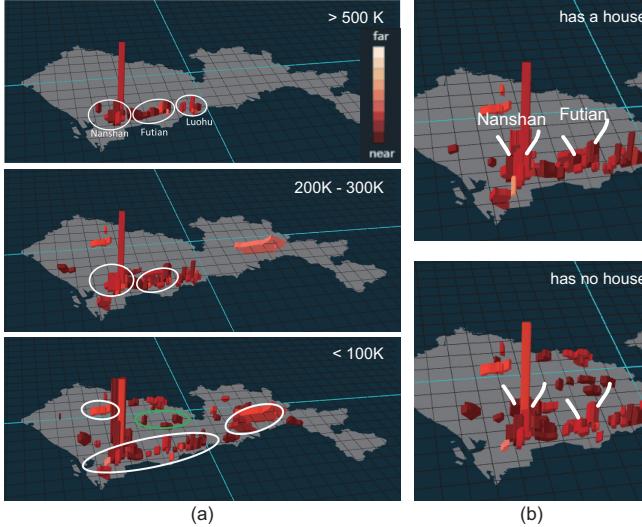


FIGURE 13: Home-work distances of different groups: (a) with different salary levels; (b) with or without house.

stands for visiting frequency of the TAZ and its color encodes the average home-work distance, red for close and yellow for far. As Fig. 14(a) shows, for all the three groups in different income levels, most of the TAZs have dark red or red color which indicates the home-work distance is generally short, especially in downtown. Only those with less than 100K salary work outside the center, such as the space in the green circle. When comparing the three groups in the downtown area, it is seen that it takes more time to go to work in Futian District as the decrease of the income.

We explore the home-work distance of groups with and without a house. Fig. 14(b) shows the close-up views. It is found that the Nanshan District has a shorter home-work distance for individuals without a house, while Futian District has shorter for those with a house. Nanshan is relatively newer to Futian so that there might be more easily-accessible houses to be rent.

C. CASE 3: SPACE DISTRIBUTION OF LIFESTYLES VS JOBS

In this case, we explore the correlation between mobility patterns with the job, to see whether the job has an impact on the movement. *Officer*, *workman*, *student*, and *businessman* are selected as examples and the population of trips for certain purposes is shown In Fig. 15. The height of bars encodes the amount of visit to each TAZ, and color encodes the traveling purpose. There are 10 different kinds of traveling purposes. For quick linking between color and purposes, the regular trips such as go to work, school and home are dyed in cold color, and the irregular trips such as go shopping, hospital, etc, are dyed in the warm color. As Fig. 15 shows, for officers, the TAZs changes dramatically in height. Some TAZs have much more visiting than others because officers are more constrained to work in some buildings than individuals with other jobs. The relative percentage of irregular activity

differs for the four groups. For Workman, about 80 percent of movements are the regular purpose, either go to work or go back home. For students, the percentage becomes to be about 50 percentage. Compared to them, officer and businessman have more flexible and diversified lifestyle.



FIGURE 14: Trip purposes distribution of four groups with different jobs

Next, we concentrated on one purpose to see whether it is related to jobs. Here we use “dinner or entertain” as the example. As Fig 16 shows, the height of TAZ bars represents the visiting amount of trips for dinner or entertainment. Color stands for moving distance to reach this TAZ. As we can see, officers prefer to entertain in the downtown area, while workmen cover a larger area. Students tend to concentrate on Nanshan district, probably near schools. For the businessmen, they go lots of places for dinner or entertainment. The biggest difference with other groups is that they would travel a long distance to go somewhere far from downtown areas.

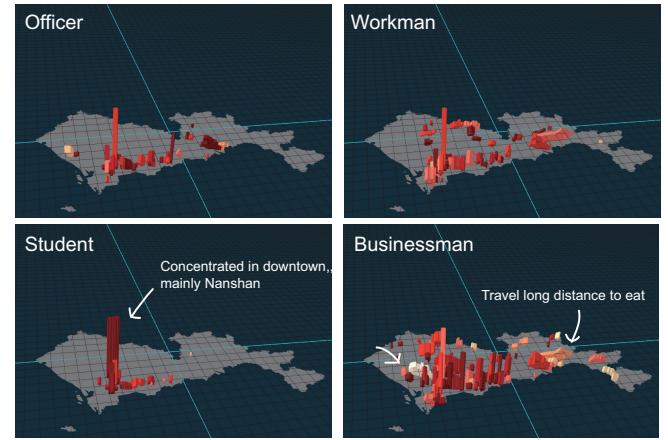


FIGURE 15: The “dinner or entertain” purpose of four groups with different jobs

VII. EXPERT FEEDBACK

We ran an interview with the two domain experts, to ask about their feedbacks on the designed system. Overall, the two domain experts praised high for the full walk-through

from data collection to analysis, towards the research problem of spatial preference with demographic characteristics. They also appreciated the overall design of a visual analytic system and the finding cases. Specifically, they pointed out what they like and also to be improved.

Expert A expressed her like in the data-driven profile visualization. She commented that “it takes little effort to learn how to read it, which really taps into my daily experience.”. Expert B also commented the data-driven profile visualization provides an intuitive representation of individuals, though sometimes it is cumbersome to recognize each of them when there are too many. But he said the combination usage of profile and t-SNE is very efficient, which enables him to quickly identify individuals from the mass. Both two experts spoke highly about dynamic identifying groups of interest and on-the-fly analysis of the spatial preference. Expert A commented that “the visual analytic system allows me to try any group of interest”. Expert B like the cross-filtering interactions across multiple views. For spatial visualization, both two experts thought our solution, 2.5D visualization is a grid-based spatial visualization is a fancy design choice. With detail slicing function, not only the overview of the spatial preference, but also the inner details of each TAZ could be filtered out and check directly.

VIII. CONCLUSION

This work demonstrates a throughout research flow from data collection, basic analysis to deep visual analysis of spatial preference with individual demographic characteristics. Considering the irresistible trend in mixing demographic data and spatial data to solve the urban problem, this work takes one step forward in exploring the relationship between movement and individuals’ characteristics, which is rarely touched by previous geo-tagged social media visual analytics. The design of the visual analytic system is driven by three tasks about individuals and mobility patterns. Tapping into our daily experiences, the data-driven profile visualization provides an organic and intuitive representation of individuals. Its effectiveness in identifying similar or abnormal individuals from the mass has been demonstrated. The design of 2.5D spatial visualization continues the exploration of the typical research problem how to visualize spatial information and abstract information together. Although it is not a one-step solution, 2.5D visualization with the detail slicing is able to provide an occlusion-free analysis with spatial context preserved. Our current research focuses on fundamental facets of the spatial preference and doesn’t draw a strong conclusion on the correlation between movements and individual characteristics. In next step, more delicate algorithms will be involved to extract high-level spatial preference, such as the frequent visiting sequence of an individual group, etc.

REFERENCES

- [1] J. Hao, “Folding beijing,” *Invisible Planets*, pp. 219–293, 2016.
- [2] R. E. Pahl, *Whose city?*, 1975.
- [3] M. Mayer, “Whose city? from ray pahl’s critique of the keynesian city to the contestations around neoliberal urbanism,” *The Sociological Review*, vol. 65, no. 2, pp. 168–183, 2017. [Online]. Available: <https://doi.org/10.1111/1467-954X.12414>
- [4] Q.-L. Gao, Q.-Q. Li, Y. Yue, Y. Zhuang, Z.-P. Chen, and H. Kong, “Exploring changes in the spatial distribution of the low-to-moderate income group using transit smart card data,” *Computers Environment and Urban Systems*, vol. 72, pp. 68–77, 2018. [Online]. Available: <Go to ISI>://WOS:000447102000007
- [5] Y. Yue, H.-d. Wang, B. Hu, Q.-q. Li, Y.-g. Li, and A. G. O. Yeh, “Exploratory calibration of a spatial interaction model using taxi gps trajectories,” *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 140–153, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0198971511000901>
- [6] Özge Öner, “Retail city: the relationship between place attractiveness and accessibility to shops,” *Spatial Economic Analysis*, vol. 12, no. 1, pp. 72–91, 2017.
- [7] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, “Visual traffic jam analysis based on trajectory data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.
- [8] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, “Visualizing interchange patterns in massive movement data,” *Computer Graphics Forum*, vol. 32, no. 3pt3, pp. 271–280, 2013.
- [9] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. Zhang, “E-map: A visual analytics approach for exploring significant event evolutions in social media,” in *Proceedings of the IEEE Conference on Visual Analytics Science&Technology (VAST)*, 2017.
- [10] H. Bosch, D. Thom, F. Heimerl, E. Püttmann, S. Koch, R. Krüger, M. Wörner, and T. Ertl, “Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2022–2031, 2013.
- [11] Y. Zheng, “Trajectory data mining: an overview,” *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 29:1–29:41, 2015.
- [12] W. Chen, F. Guo, and F.-Y. Wang, “A survey of traffic data visualization,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2970–2984, 2015.
- [13] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, “TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection,” in *Proceedings of IEEE Pacific Visualization Symposium (PacificVis)*, 2011, pp. 163–170.
- [14] M. Lu, Z. Wang, and X. Yuan, “Trajrank: Exploring travel behaviour on a route by trajectory ranking,” in *Visualization Symposium (PacificVis)*, 2015 IEEE Pacific. IEEE, 2015, pp. 311–318.
- [15] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni, “Visual analysis of route diversity,” in *IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2011, pp. 171–180.
- [16] M. Lu, C. Lai, T. Ye, J. Liang, and X. Yuan, “Visual analysis of multiple route choices based on general gps trajectories,” *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 234–247, June 2017.
- [17] Y. Liu, X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi, “Social sensing: A new approach to understanding our socioeconomic environments,” *Annals of the Association of American Geographers*, vol. 105, no. 3, pp. 512–530, 2015.
- [18] 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 Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2649–2658, 2012.
- [19] R. Krueger, D. Thom, and T. Ertl, “Visual analysis of movement behavior using web data for context enrichment,” in *Visualization Symposium (PacificVis)*, 2014 IEEE Pacific. IEEE, 2014, pp. 193–200.
- [20] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. L. Zhang, and J. Zhang, “Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data,” *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 270–279, 2016.
- [21] S. T. Peddinti, K. W. Ross, and J. Cappos, “On the internet, nobody knows you’re a dog: A twitter case study of anonymity in social networks,” in *Proceedings of the second ACM conference on Online social networks*. ACM, 2014, pp. 83–94.
- [22] F. Luo, G. Cao, K. Mulligan, and X. Li, “Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of chicago,” *Applied Geography*, vol. 70, pp. 11–25, 2016.
- [23] P. A. Longley, M. Adnan, and G. Lansley, “The geotemporal demographics of twitter usage,” *Environment and Planning A: Economy*

- and Space, vol. 47, no. 2, pp. 465–484, 2015. [Online]. Available: <https://doi.org/10.1068/a130122p>
- [24] P. A. Longley and M. Adnan, “Geo-temporal twitter demographics,” International Journal of Geographical Information Science, vol. 30, no. 2, pp. 369–389, 2016. [Online]. Available: <https://doi.org/10.1080/13658816.2015.1089441>
- [25] M. Batty, The new science of cities. Mit Press, 2013.
- [26] B. Shneiderman, “The eyes have it: a task by data type taxonomy for information visualizations,” in Proceedings 1996 IEEE Symposium on Visual Languages, Conference Proceedings, pp. 336–343.
- [27] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, “Glyph-based visualization: Foundations, design guidelines, techniques and applications.” in Eurographics (STARs), 2013, pp. 39–63.
- [28] H. Chernoff, “The use of faces to represent points in k-dimensional space graphically,” Journal of the American Statistical Association, vol. 68, no. 342, pp. 361–368, 1973.
- [29] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” Journal of machine learning research, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [30] C. Weaver, “Cross-filtered views for multidimensional visual analysis,” IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 2, pp. 192–204, 2010.
- [31] G. Sun, R. Liang, H. Qu, and Y. Wu, “Embedding spatio-temporal information into maps by route-zooming,” IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 5, pp. 1506–1519, 2017.
- [32] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, “Stacking-based visualization of trajectory attribute data,” IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 12, pp. 2565–2574, 2012.



QINGQUAN LI received the Ph.D. degree in geographic information system (GIS) and photogrammetry from Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1998. He is currently a Professor with Shenzhen University, Guangdong, China, and Wuhan University, Wuhan, China. His research areas include 3-D and dynamic data modeling in GIS, location-based service, surveying engineering, integration of GIS, Global Positioning System and remote sensing, intelligent transportation system, and urban informatics.

• • •



QI WANG received the B.Eng. degree in information security from Wuhan University, China. She is currently working toward the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. Her research interests include transport geographic information system, spatiotemporal data analysis, mining and visualization.



MIN LU is an associate researcher at ShenZhen University. Her major research interest is data visualization and visual analytics, especially focusing on the urban data. She received Ph.D degree in Computer Science from Peking University in 2017 and BS degree in Computer Science from Beijing Normal University in 2011.