

Unfolding the City: Exploring Urban Spatial Attractiveness with Social Characteristics

Qi Wang · Min Lu · Yang Yue · Qingquan Li

Received: date / Accepted: date

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 characteristics visit places with different weights and preferences, sculpturing the city into a space of diverse mobility patterns. Different from the well study of anonymous travel behaviors, the mobility patterns of characterized individuals, is less studied because of 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 develop a visual analytic system to investigate the mobility patterns of groups by specifying 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.

Keywords Visual analytics · mobility pattern · individual characteristics · city

1 Introduction

[WQ: change mobility pattern to spatial attractiveness] The science fiction Folding Beijing (Hao, 2016) depicts a world where three classes of people live in independent

Q. Wang · Q. Li()

the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, and also ShenZhen Key Laboratory of Spatial Smart Sensing and Services, ShenZhen University, Shenzhen 518060, China

Tel.: +86-13297983047

E-mail: wangqi@whu.edu.cn, liqq@szu.edu.cn

M. Lu · Y. Yue

School of Architecture and Urban Planning, and also ShenZhen Key Laboratory of Spatial Smart Sensing and Services, ShenZhen University, Shenzhen 518060, China

E-mail: minlu@szu.edu.cn, yueyang@szu.edu.cn

spatial-temporal patterns, though sharing the same earth surface. This is an artistical metaphor for one potential fact that people make different decisions in visiting places according to individual determinants, such as social, economic, educational factors, etc. This phenomenon has attracted the research interest for a long time in the sociological field. Back to 1970, Pahl (Pahl, 1975) posed the question ‘whose city’ and interest in understanding the territorial inequalities. Contemporary contestation also continually raise the banner ‘The City Belongs to All!’ (Mayer, 2017). Having a good understanding of how the mobility patterns relate to characterized individuals sheds light on region utilization, diversity mix-up, and opportunities promotion.

In the past decades, the power of visual analytics in exploring spatial-temporal data has been recognized and well developed. Transportation problems are answered by various movement data, such as traffic jams in taxi GPS trajectoires (Wang et al., 2013), interchanging flows in public transportation (Zeng et al., 2013), etc. 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 (Chen et al., 2017) or topics (Bosch et al., 2013). Those works essentially contribute in deepening the contextual understanding of moving activity. But they rarely combine the mobility with individuals’ characteristics because which cannot be 100% inferred from social media profiles. There is still a research gap in exploring the relationship between mobility pattern and characteristics.

In this work, we contextualize the analysis of mobility pattern in the backcloth of individual characteristics. Our motivation is to fuse the movement data with individuals’ profiles in an attempt to ascertain the dynamics pattern of groups with different characteristics. To this, we need to get the characterized information of movers. The sensitivity issue in privacy data is one of the main challenges. 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 count information online. In this work, with a census survey released via social media application, we ran an online demographical collection campaign. Then we developed a visual analytics system to explore the mobility patterns centered around people. 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 mobility patterns. Finally, applying our method in Shenzhen, several cases are derived to demonstrate the powers of our method. There are two contributions of this work:

- (1) investigate mobility patterns with 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.

2 Related work

This work concerns the research topic of spatial data visual analytics. Spatio-temporal data has three components, i.e., spatial, temporal and thematic (Andrienko and Andrienko, 2013). Lots of related work focus on the spatio-temporal 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 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.

2.1 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 (Zheng, 2015). 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 (Chen et al., 2015). A large number of visual analytics tools and applications cover situation-aware exploration, pattern discovery and traffic situation monitoring. Wang et al (Wang et al., 2013) extract the traffic jam propagation graph extraction to reveal underlying data patterns. Guo et al. (Guo et al., 2011) and Zeng et al. (Zeng et al., 2013) construct geographical regions and visually aggregate the in-between movements as flows. To discover the route travel patterns, Lu et al. (Lu et al., 2015) propose TrajRank to explore the route travel behavior based on ranking. For multiple routes, Liu et al. (Liu et al., 2011) study the route diversity between locations and Lu et al. (Lu et al., 2017) explore the route choice behavior among multiple routes.

2.2 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 (Liu et al., 2015). 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. (Cao et al., 2012) propose Whisper for tracing the pathways of tweets on a spatial hierarchical layout, to investigate how information flow among multiple places. Krueger et al. (Krueger et al., 2014) used GPS and location-based service data to support the analysis of movement behaviors. Chen et al. (Chen et al., 2016) 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 (Peddinti et al., 2014), to break down the demographic characteristics of social media users. Luo et al. (Luo et al., 2016) 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. (Longley et al., 2015) (Longley and Adnan, 2016) identifies and assess the biases inherent in social media usage in social research and evaluate the deployment of social media data in research applications.

3 Online Census

The advent of mobile sensing techniques and social media applications makes it possible to collect spatial data from the social media source. Complementary to the conventional census, it brings the benefit of larger sampling frequency and a broader range in terms of space and time. It is possible to reach a wide range of individuals and collect the movement in human inactive time, such as the mid-night.



Fig. 1 Census Interface: (a) personal characteristics collecting page; (b) trips collecting page; (c) credit system page

In this work, we perform the census survey in Shenzhen, which is one of the most modern metropolia in China. The experiment is deployed on Wechat, a widely used social media application. Figure 1 shows the data collecting interfaces. Each individual hands in his or her personal characteristics. For privacy issue, all detailed personal information are desensitized to categorical levels (Figure 1(a)).

Individual Characteristics Figure 2 lists the eight domains, including social, economic and demographic aspects, to give a generalized description of the individual characteristics. The profile will serve as the ingredients for the analysis of mobility patterns over diverse individuals.

Traveling Trips Besides to those individual characteristics, each individual can upload dynamic traveling trips (Figure 1(b)). Each trip requires the information of start/end location, start/end time, traveling purpose. To encourage the trip uploading, a credit system retains the contribution of individuals on trips and rewards the volunteers with the top credits (Figure 1(c)).

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

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

3.1 Basic Statistics of Data

Over the releasing time period from 2015-11 to 2016-01, 21435 individuals (48% females and 52% males) were reached and 229155 trips are collected. Each volunteer contributes 11 trips average.

Our case-study data is confined to a small proportion of Wechat users who opt to contribute their information and trips. Considering the caveat that self-selecting individuals are most unlikely to represent any clearly defined population (Longley et al., 2015), we performed a series of preliminary statistics to check whether it is rich enough to represent a wide range of the population in the city.

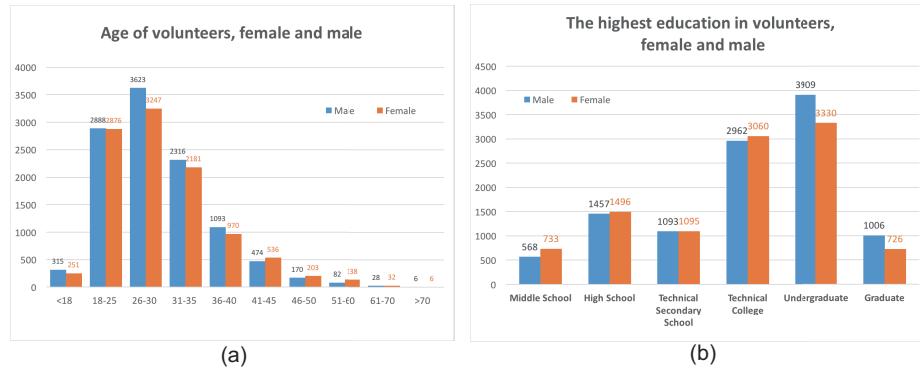


Fig. 3 Age and Education Distribution: (a) age; (b) education

Figure 3 gives the distribution of age and education over the population. It shows that samples cover a wide range of ages, dominating between 18 to 45. There is also a

few records pertaining to individuals below the age of 18 or above 70. The distribution follows the fact that Shenzhen is a city where the majority is young people. According to the 2015 Annual Census Statistics report¹, people aging 15-64 occupy 83.23% and the median age is 31.5. Figure 3 gives the distribution of education levels, ranging from low to high. The technical college and university dominate the samples at the 61% occupancy rate.

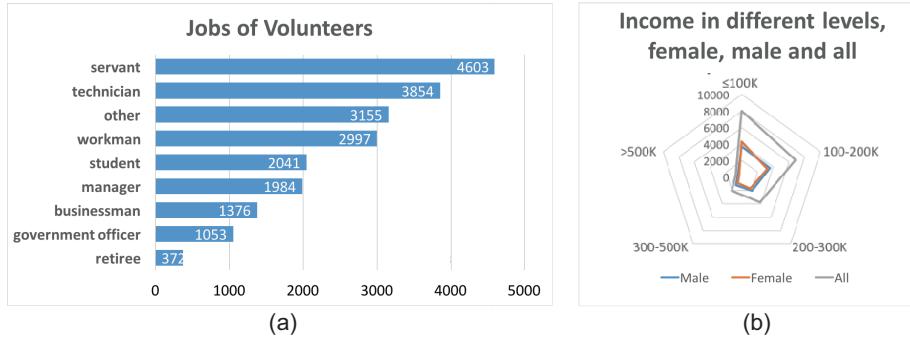


Fig. 4 Job and Income Distribution: (a) job; (b) income

Figure 4(a) shows the job types of sampled individuals, who are servants, workers, officers, businessmen and so on. The covering of jobs is pretty wide. Figure 4(b) gives the radar diagram of the annual pay. The majority get paid below 200 000. Individuals with higher salary are also reached in our census.

Figure 5 shows the basic statistics related to trips. In Figure 5(a), the active traveling time (here, we take the start time as the representative of active time) follows the common knowledge of urban life. There are obvious morning and late afternoon peaks. Figure 5(b) gives the counting of different traveling purposes. 95% trips are tagged with clear raveling 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. Figure 5(c) 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 (Batty, 2013).

Because there is always inevitable bias inherent in fully representing the ground truth of the population, the preliminary statistical analysis shows positive sign of a relatively even sampling of the population.

4 Overview

[WQ: Adding a background. Expert discussion. Summarize the tasks below.-0419]
[LM: Add a paragraph to introduce the motivation of this work, where we derive the tasks via the interview with domain experts.-0420]

¹ http://www.szstj.gov.cn/xxgk/tjsj/pcgb/201606/t20160614_3697000.htm

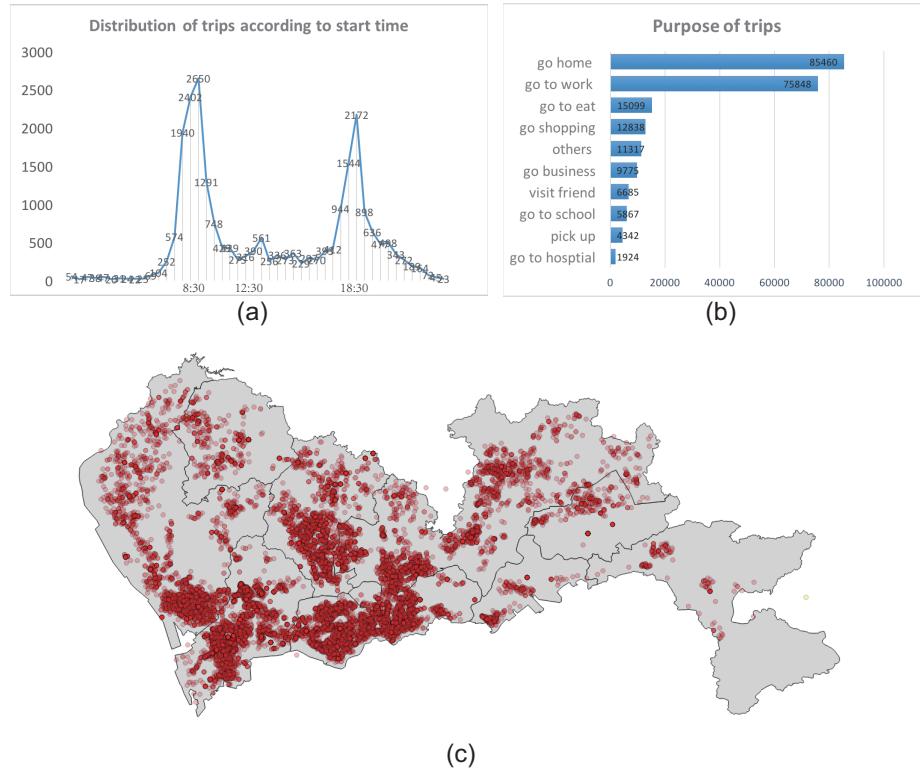


Fig. 5 Statistics of Trips: (a) start time of trips; (b) purpose of trips; (c) distribution of origins/destinations

[WQ: Common human mobility indicators: travel distance, travel time, travel frequency, radius of gyration and movement entropy.]

Spatial attractiveness is an important concept in GIS. Explain why. To derive the details, we interview the expert in GIS. Some questions like, how citizens live in the city, where they live, work and entertain? Which place is popular? For different citizens, whether a place has the same attraction? ... With these questions, we summarize them into the tasks below, which need to be answered by our system.

The design of the system is driven by tasks, to align the analysis of mobility patterns with individual characteristics. Before diving into the design of the system, introduce the couple of specific tasks the system intended for.

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

4.1 Design Considerations

With these three tasks, we derive following design considerations:

- Intuitive perception of an individual as an organic complex (C1): the system should make use of users' daily life experience in knowing people 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 mantra by Shneiderman (Shneiderman, 1996), 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 system should maintain flexible interactions which allow the end-users to explore freely.

5 Visual Design

With the design consideration (introduced in Section ??), we design a visual analytic system... Show a overview figure, to introduce the relationship between multiple views

With those design considerations, we develop a visual analytics system as Figure 6 shows. [WQ: instead of fig. teaser] 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 chosen group of individuals, its moving related information is visualized and explored in a 2.5D spatio-temporal view.

5.1 Data-driven Profile Visualization

Glyph-based visualization Borgo et al. (2013) 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. Dual glyphs are designed, to illustrate the information of individual from two different perspectives, i.e., detailed inspection of individuals and efficient comparsion over multiple individuals.

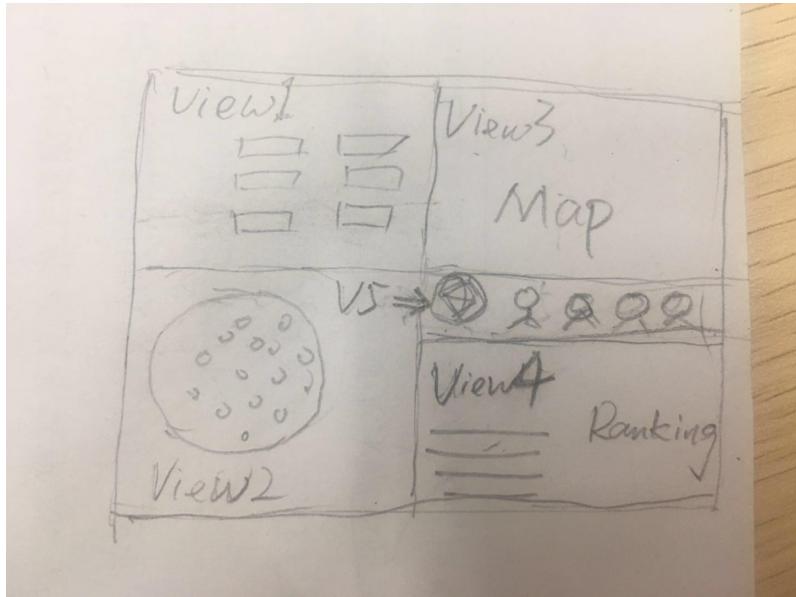


Fig. 6 System Interface: (a) an interactive infographics provides the basic statistic facets of individuals; (b) View1: a XXXX visualization gives an overview of individuals, where the similarity in high dimensional space is preserved in 2D space; (c) View2: data-driven profiles support users with organic individual perception; (d) View3: map view for movement of the chosen group; (e) View4: for ranking or comparison; (g) View5: top N kinds of citizens with social characteristics.

Star-like Glyph To compare over multiple individuals, a star-like glyph is designed to summary each individual, to ease the comparison by common cooordinating configurations....

Cartoon Glyph ... Among glyph design, Chernoff Face Chernoff (1973) 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 (C1). 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.

Figure ?? 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 the gender 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 gray. 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.

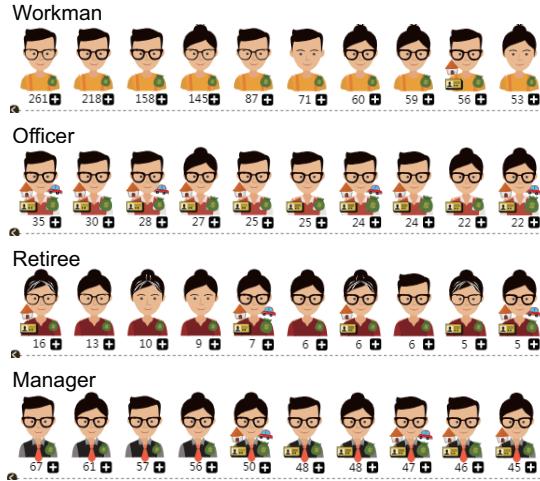


Fig. 7 Representative figures with the top 10 largest population in four jobs

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. Figure 7 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.

5.2 XXXX Projection of Individuals

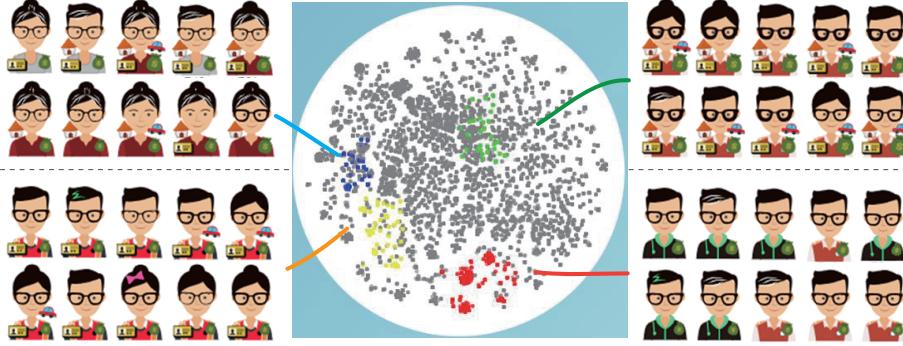


Fig. 8 t-SNE project with four groups of the interest: (a) t-SNE project

[WQ: A octagon radar graph to position all citizens. A clustering algorithm for command similar persons into a new group.(maybe: interactions to set the weight of each characteristic)]

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 (Maaten and Hinton, 2008) 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 Figure 8 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. Figure 8 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 machinesm (Weaver, 2010). 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 characteristical variables.

5.3 2.5D Spatial Visualization

[WQ: this part may be deleted.]

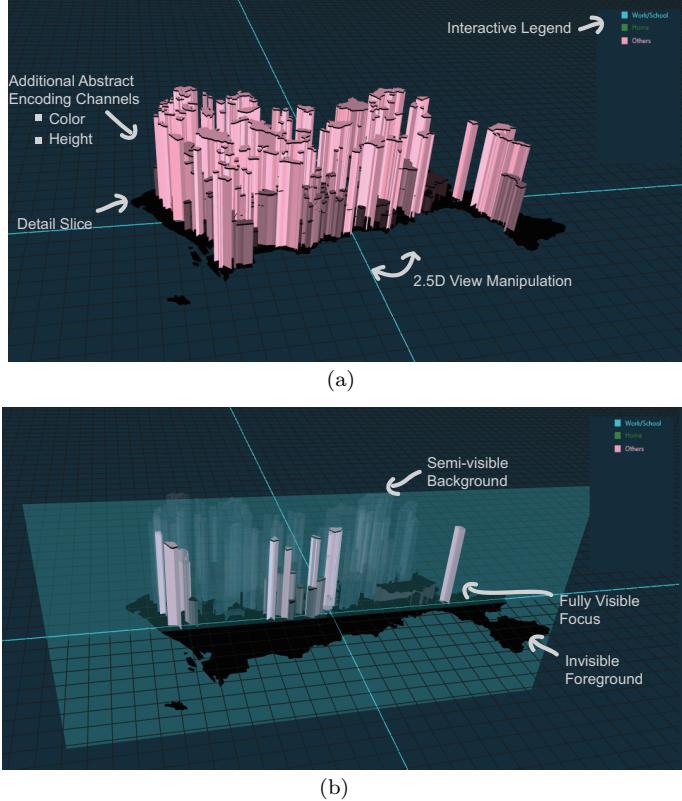


Fig. 9 2.5D Spatial Visualization: (a) 2.5D spatial visualization with z-axis to encode addition information; (b) detail slicing in 2.5D space.

Embedding multiple variables in the spatial map is a challenging problem. Distorting 2D spatial for better space usage, such as the partial route embedding (Sun et al., 2017) 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 (Tomin-ski et al., 2012), the space in this work is embedded in 2.5D space, to relax the z-axis for additional abstract information encoding.

Figure 9(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 Figure 9(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 are so continuously flexible that users probably find a suitable observing point. In the cases where occlusion is inevitable, an interaction named as Detail Slicing is developed, as Figure 9(b) shows. 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 (Figure ??(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.

5.4 A glyph design details (in view3)

We design a kind of glyph to express a high-dimensional travel indicators. It will be embeded in the map view[WQ: for case1].

5.5 Ranking or comparison method (in view4)

[WQ: for case3]We design ranking/comparison method with different weight of different purpose or different groups.

6 Experts Feedback

[LM: Add a section to introduce the domain experts' feedback]

We interview XX domain experts from the GIS field. XX of them are with ...
The procedure went as following. We first introduce them the system an...

7 Case study

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.

7.1 Case 1: Work-life Balance VS Income/Education

[WQ: Aim1: use view1 and view2 for groups. Aim2: show their basic activity spaces.]

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.

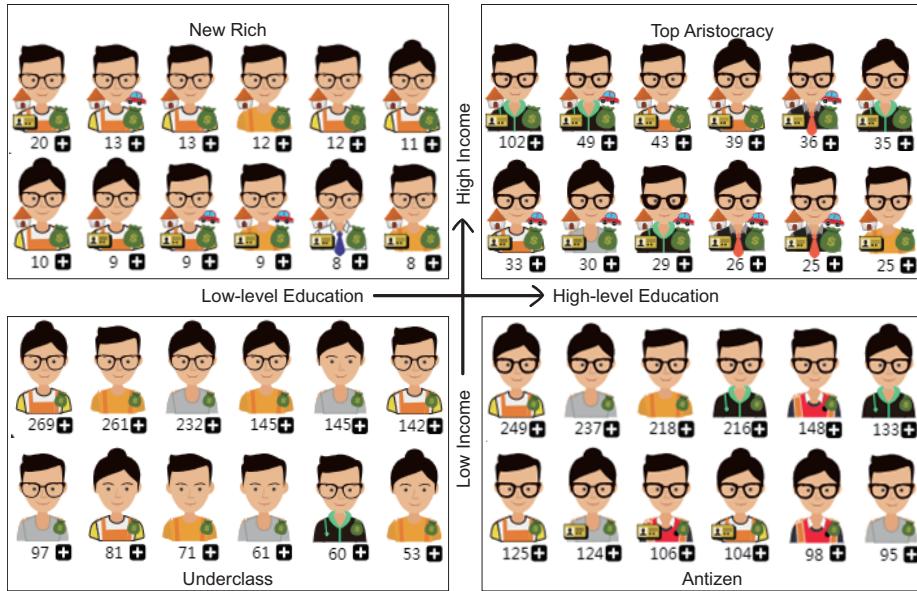


Fig. 10 Four Groups Divided by Income and Education Characteristics

Although it is incomplete, income and education are two of the most essential characteristics to tag an individual with a good resource or not.

EDIT paragraph below. not just give number, use view1 and view2 to filter the target groups.

As Figure 10 shows, taking income and education as the two dimensions, taking 300K 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 groups 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.

Figure 10 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 Figure 11 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 antizen live in Nanshan

District, where there are many universities and high-tech industrial parks full of well educated people. The glyphs on the map means the travel distance(even differentiate home-work distance(which is an important in mobility research), and extended travel distance), travel radius. [WQ: the 3rd design in Section.design, a kind of glyph for high-dimensional mobility indicator]

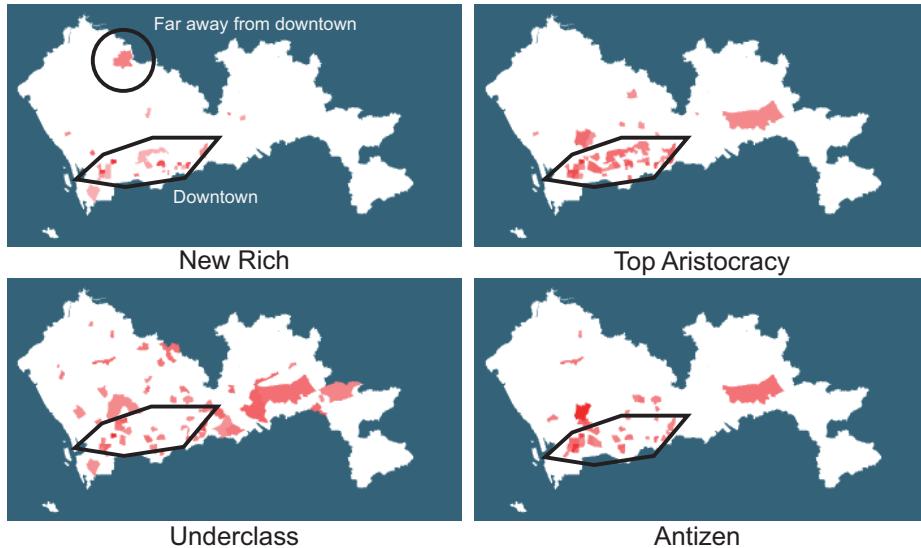
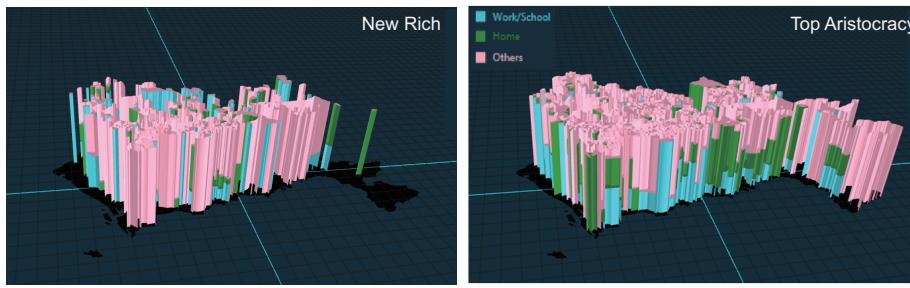


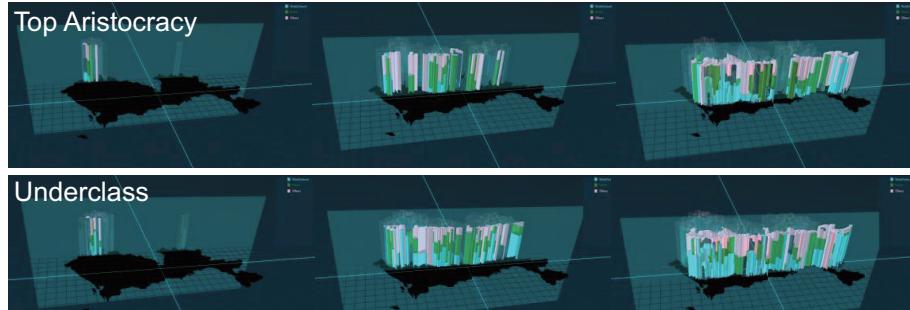
Fig. 11 Home Distributions of the Four Groups

[WQ: paragraph below may be deleted. It is about attractiveness entropy, which in case3] 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. Figure 12(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 Figure 12, 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 5.3) is applied to the four groups. Figure 12(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.

7.2 Case 2: Spatial Attractiveness Based on One Purpose



(a)



(b)

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

[WQ: aim: select one purpose; View3 shows spatial attractiveness, including: quantity, distance, direction ; View4 shows ranking of origins, (or others); (add a interaction in view5 for citizens in selected grid.)]

Due to the various functions of the areas, for different space, the attractiveness for different aims differs. Therefore, in this case, we demonstrate the use of the system for exploring the spatial attractiveness for one purpose.

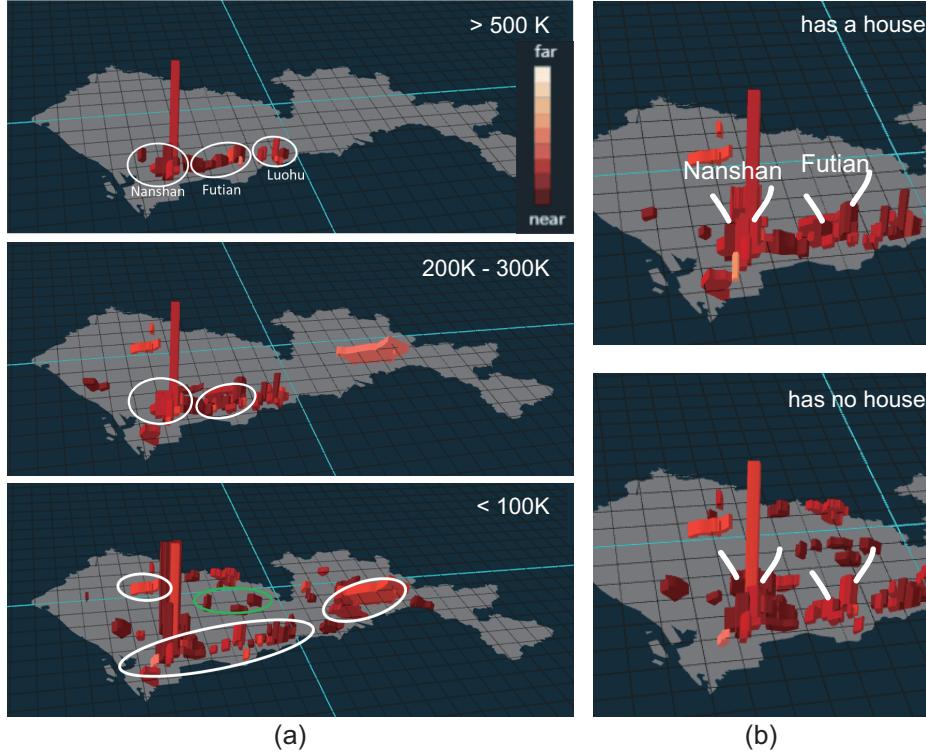


Fig. 13 Home-work distances of different groups: (a) with different salary levels; (b) with or without house.

Home-work distance is an essential metric in economics, geography, and sociology. It refers to the distance between house and workplace. 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 Figure 13, the height of the bar stands for visiting frequency of the TAZ and its color encodes the average home-work distance, red for close and yellow for far. As Figure 13(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. Figure 13(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.

7.3 Case 3: Spatial Attractiveness Based on Mixed Purposes

[WQ: aim: with mixed purposes, show attractiveness quantity and entropy. View3 shows spatial attractiveness, including: quantity, distance, direction ; View4 shows ranking of origins, (or others); (add a interaction in view5 for citizens in selected grid.)]

Due to the various functions of the areas, for different space, the attractiveness for different aims differs. Therefore, in this case, we demonstrate the use of the system for exploring the spatial attractiveness for one purpose.

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 Figure 14. 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 Figure 14 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.

Next, we concentrated on one purpose to see whether it is related to jobs. Here we use “dinner or entertainment” as the example. As Fig 15 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.

8 Conclusion

This work demonstrates a throughout research flow from data collection, basic analysis to deep visual analysis of mobility patterns with individual characteristics. Considering the irresistible trend in mixing social data and spatial data to solve the

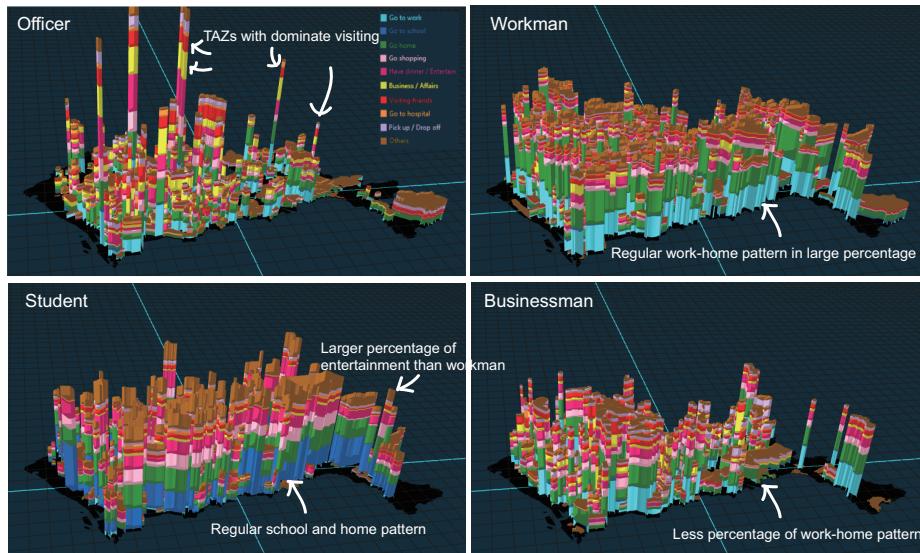


Fig. 14 Trip purposes distribution of four groups with different jobs

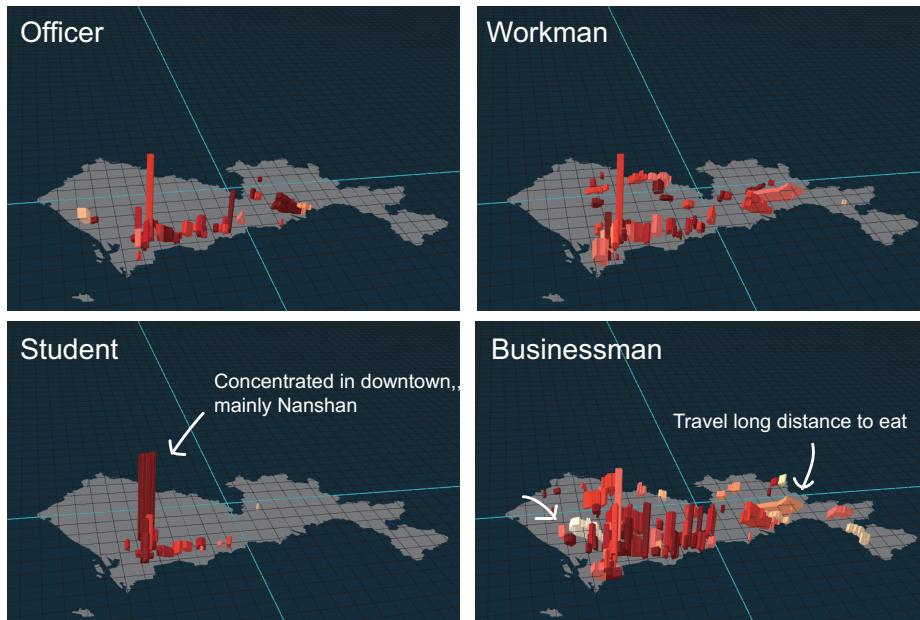


Fig. 15 The “dinner or entertain” purpose of four groups with different jobs

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 intu-

itive 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 mobility patterns and doesn't draw a strong conclusion on the correlation between mobility pattern and individual characteristics. In next step, more delicate algorithms will be involved to extract high-level mobility patterns, such as the frequent visiting sequence of an individual group, etc. On the other hand, the data we used in this work shows its potential in representing the population. As the next step, research effort can be put into a more serious verification of how small sampled data can represent a large population.

References

- Andrienko N, Andrienko G (2013) A visual analytics framework for spatio-temporal analysis and modelling. *Data Mining and Knowledge Discovery* pp 1–29
- Batty M (2013) The new science of cities. Mit Press
- Borgo R, Kehrer J, Chung DH, Maguire E, Laramee RS, Hauser H, Ward M, Chen M (2013) Glyph-based visualization: Foundations, design guidelines, techniques and applications. In: Eurographics (STARs), pp 39–63
- Bosch H, Thom D, Heimerl F, Püttmann E, Koch S, Krüger R, Wörner M, Ertl T (2013) Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics* 19(12):2022–2031
- Cao N, Lin YR, Sun X, Lazer D, Liu S, Qu H (2012) Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics* 18(12):2649–2658
- Chen S, Yuan X, Wang Z, Guo C, Liang J, Wang Z, Zhang XL, Zhang J (2016) Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics* 22(1):270–279
- Chen S, Chen S, Lin L, Yuan X, Liang J, Zhang X (2017) 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)
- Chen W, Guo F, Wang FY (2015) A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems* 16(6):2970–2984
- Chernoff H (1973) The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association* 68(342):361–368
- Guo H, Wang Z, Yu B, Zhao H, Yuan X (2011) 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), pp 163–170
- Hao J (2016) Folding beijing. *Invisible Planets* pp 219–293

- Krueger R, Thom D, Ertl T (2014) Visual analysis of movement behavior using web data for context enrichment. In: Visualization Symposium (PacificVis), 2014 IEEE Pacific, IEEE, pp 193–200
- Liu H, Gao Y, Lu L, Liu S, Qu H, Ni LM (2011) Visual analysis of route diversity. In: IEEE Conference on Visual Analytics Science and Technology (VAST), IEEE, pp 171–180
- Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, Chi G, Shi L (2015) Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers* 105(3):512–530
- Longley PA, Adnan M (2016) Geo-temporal twitter demographics. *International Journal of Geographical Information Science* 30(2):369–389, DOI 10.1080/13658816.2015.1089441, URL <https://doi.org/10.1080/13658816.2015.1089441>
- Longley PA, Adnan M, Lansley G (2015) The geotemporal demographics of twitter usage. *Environment and Planning A: Economy and Space* 47(2):465–484, DOI 10.1068/a130122p, URL <https://doi.org/10.1068/a130122p>, <https://doi.org/10.1068/a130122p>
- Lu M, Wang Z, Yuan X (2015) Trajrank: Exploring travel behaviour on a route by trajectory ranking. In: Visualization Symposium (PacificVis), 2015 IEEE Pacific, IEEE, pp 311–318
- Lu M, Lai C, Ye T, Liang J, Yuan X (2017) Visual analysis of multiple route choices based on general gps trajectories. *IEEE Transactions on Big Data* 3(2):234–247, DOI 10.1109/TBDA.2017.2667700
- Luo F, Cao G, Mulligan K, Li X (2016) Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of chicago. *Applied Geography* 70:11–25
- Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9(Nov):2579–2605
- Mayer M (2017) Whose city? from ray pahlâ s critique of the keynesian city to the contestations around neoliberal urbanism. *The Sociological Review* 65(2):168–183, DOI 10.1111/1467-954X.12414, URL <https://doi.org/10.1111/1467-954X.12414>, <https://doi.org/10.1111/1467-954X.12414>
- Pahl RE (1975) Whose city?
- Peddinti ST, Ross KW, Cappos J (2014) 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, pp 83–94
- Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE Symposium on Visual Languages, pp 336–343, DOI 10.1109/VL.1996.545307
- Sun G, Liang R, Qu H, Wu Y (2017) Embedding spatio-temporal information into maps by route-zooming. *IEEE Transactions on Visualization and Computer Graphics* 23(5):1506–1519
- Tominski C, Schumann H, Andrienko G, Andrienko N (2012) Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics* 18(12):2565–2574
- Wang Z, Lu M, Yuan X, Zhang J, Van De Wetering H (2013) Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics* 19(12):2159–2168

-
- Weaver C (2010) Cross-filtered views for multidimensional visual analysis. *IEEE Transactions on Visualization and Computer Graphics* 16(2):192–204
- Zeng W, Fu CW, Arisona SM, Qu H (2013) Visualizing interchange patterns in massive movement data. *Computer Graphics Forum* 32(3pt3):271–280
- Zheng Y (2015) Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology* 6(3):29:1–29:41, DOI 10.1145/2743025