Technische Universität München<br>Department of Civil, Geo and Environmental Engineering<br>Chair of Cartography<br>Prof. Dr.-Ing. Liqiu Meng

# Building visual overview of potential inefficiencies in heterogeneous mobility system 

## Yueying Lu

Master's thesis

Duration: 01.05.2017-22.12.2017
Study Course: Cartography M.Sc.
Supervisor: Dr.-Ing Mathias Jahnke
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner
Dr.-Ing Georg Fuchs
Cooperation: Fraunhofer IAIS, Sankt Augustin

## Statement of Authorship

I confirm that this master's thesis is my own work and I have documented all sources and material used.

## Acknowledgement

Foremost, I would like to express my sincere gratitude to my first supervisor, Dr-Ing Mathias Jahnke, for his continuous support during the whole process, insightful comments on my thesis and patience. This thesis would have never been accomplished without his guidance.

I would like to thank Dr. Georg Fuchs, my supervisor from Fraunhofer IAIS, for giving me the opportunity to work on this exciting project and for his guidance during research. I am deeply grateful to Drs. Gennady Andrienko and Natalia Andrienko from Faunhofer IAIS, for giving constructive comments and patient help. I would like to show my greatest appreciation to Dr. Siming Chen from Fraunhofer IAIS, for helping me improve my thesis. I would like to thank M.Sc. Fabian Patterson from Fraunhofer IAIS for helping me getting and understanding data. My sincere appreciation goes to also Prof. Georg Gartner for giving me constructive advices for my thesis proposal.

I thank M.Sc. Juliane Cron for her invaluable encouragement and support when I was choosing thesis topic. In addition, I thank her as the coordinator of Cartography Master program for her excellent organizations and patient help. I thank also all the lecturers and classmates of Cartography Master program, for sharing knowledge and all valuable moments.

Finally, I owe my deepest gratitude to my family, for all their love and supporting me all the time.


#### Abstract

Finding and understanding inefficiencies in public transport is important for city planners to improve quality of public transport. However, it is challenging to understand what facts make the public transport inefficient, as well as their relationship. Especially when public transport is with multiple modes, which is defined as the heterogeneous mobility system, inefficiency is more complex. Visualization and visual analytics approach, which combines the human knowledge and automatic computation, provides good opportunities to understand such not welldefined heterogeneous mobility system inefficiency patterns. In this thesis, we first review the state of the art research works in transportation analysis and visualization. By extending the existing literatures, we propose a summary of inefficiency features with detailed analysis and description, including unreliability, frequency, complexity and accessibility of heterogeneous mobility system. Unreliability indicates the delay and advance of different transportation. Frequency describes the different length of headways in multiple transportation methods. Complexity addresses on the multiple times of transfer while the accessibility summaries the reachability from one origin to the city regions. These features work together and affect the inefficiency in transportation system. We use visual analytics methods to enable target users (e.g. city planner) understand these inefficiency features with proper visual encodings. Our case study of Warsaw city transportation data confirms the capability of visual analytics approach to identify the inefficiencies in different perspectives of space and time.


Keywords: inefficiency, heterogeneous mobility system, visual analytics, unreliability, frequency, complexity, accessibility

## Table of Content

Statement of Authorship ..... 2
Acknowledgement ..... 3
Abstract ..... 4
List of Figures ..... 7
List of Tables ..... 10

1. Introduction ..... 11
1.1 Background ..... 11
1.2 Research Objectives ..... 12
1.3 Structure of thesis ..... 13
2. State of the art ..... 14
2.1 Key factors for efficient heterogeneous mobility system ..... 14
2.1.1 Time-related factors ..... 16
2.1.2 Access-related factors ..... 19
2.2 Visual analytics of movement data ..... 20
2.2.1 Movement visualization over space ..... 20
2.2.2 Visualization of behavior along one trajectory over time ..... 21
2.2.3 Movement events extraction ..... 21
3. Methodology ..... 22
3.1 Unreliability ..... 22
3.2 Frequency ..... 28
3.3 Complexity and accessibility ..... 29
4. Case study ..... 32
4.1 Data ..... 32
4.1.1 Data resource ..... 32
4.1.2 Data processing ..... 34
4.2 Visualization of unreliability ..... 38
4.3 Visualization of frequency ..... 49
5. Discussion ..... 54
6. Conclusion ..... 56
Reference ..... 58

## List of Figures

Figure 1 Workflow of methodology ..... 22
Figure 2 An example of trajectory wall showing speed of vehicle along thetrajectory24
Figure 3 Diverging color scheme example (Brewer et al. 1997) ..... 25
Figure 4 An example of visualization. Areas where tram 10 runs through arecolored according to excepted value of speed when it passes. The $20^{\text {th }}, 40^{\text {th }}$,$60^{\text {th }}, 80^{\text {th }}$ percentiles of the speed are colored by diverging scheme colors. 27
Figure 5 An example of travel with transfer ..... 30
Figure 6 Example of visualization of complexity and accessibility from given origin. Green circle shows waiting time at station, line shows times of transfer, blue area shows accessible area within certain time...................... 31

Figure 7 Transport stops in Warsaw city. Blue points are original stops in Warsaw and red points are centroids of a group of points inside an voronoi polygon.

Figure 8 Comparison between timetable data and real movement data. Trajectories of tram 10 on 01.09 .2016 are shown in space time cube, timetable ones are in orange and real movement ones are in blue. One selected trajectory is shown in black.35

Figure 9 Comparison between trajectory matching results. a1. original tram 10 trajectories of timetable. b1. point-based result of tram 10 b2. c. trajectorybased result of tram 10. a2 b2 c2. Corresponding trajectories shown in space time cube and black line is one selected trajectory.37

Figure 10 Enlarged comparison between trajectory matching results. a. original tram 10 trajectories of timetable. b. point-based result of tram 10 . c. trajectory-based result of tram 10. Black line is one selected trajectory. In
point-based result, it happens that at one time there are more than one position in a trajectory, which are marked in red square.

Figure 11 Trajectory wall of absolute time difference of tram 10....................... 39
Figure 12 Trajectory of interest is stroked by white lines. Significant delay is marked in blue square. Temporal information can be seen at the left side. a is at $08: 58$. b is at $17: 34$

Figure 13 Perspective-shifted trajectory wall. By shifting perspective we can see position on the real map ............................................................................. 41

Figure 14 Trajectory wall of relative time difference 42

Figure 15 Enlarged trajectory wall of relative time difference. Grey arrows on trajectory band indicate the direction of trajectory. In blue marked square, the same place within close time range, different directions show opposite relative time difference, which red is positive while green is negative. ...... 42

Figure 16 Trajectory wall with directions. a. Trajectory wall of tram 10 of direction to OsGczewska. b. Trajectory wall of tram 10 of direnction to Wycigi. In blue marked square, delay increase or decrease happens repetitively over a period.43

Figure 17 Delay comparison. Delay of a. tram during 8:30-9:30 b. bus during 8:30-9:30 c. tram during 17:00-18:00 d. bus during 17:00-18:00 e. tram of the whole day f . bus of the whole day. Legend in a applies to also $\mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}$, and f

Figure 18 Delay increase comparison. Delay increase of a. tram during 8:30-9:30 b. bus during 8:30-9:30 c. tram during 17:00-18:00 d. bus during 17:0018:00 e. tram of the whole day f. bus of the whole day. Legend in a applies to also b ,c, d, e and f ................................................................................. 47

Figure 19 Summary of delay and delay increase. a. summary result of delay b. summary result of delay increase48
Figure 20 Frequency during 07:00-19:00 on 2016-09-01 ..... 50

Figure 21 Frequency during 22:00 on 2016-09-01 to 5:00 next day ................... 50
Figure 22 Mean waiting time of high frequency trams during 13:30-14:30........ 51
Figure 23 Combination of tram frequency and delay ......................................... 52

Figure 24 Combination of bus frequency and delay........................................... 53

## List of Tables

Table 1 Definitions of public transport service quality attributes (Redman et al.,2013)15
Table 2 Description of data ..... 33

## 1.Introduction

### 1.1 Background

Cities are facing much pressure over problems caused by high car-usage. Car dependency exacerbates not only environmental concerns, such as air pollution, excessive rate of consumption of resources, but also social problems such as traffic congestion and poor health (Greene \& Wegener, 1997). One efficient solution is to reduce private car use and increase public transport use (Goodwin, 1996). With the infrastructure development of the city, heterogeneous mobility system, which is the transport system with multiple transport modes, such as train, bus, subway, etc. has been applied widely to in order to solve transportation problems and improve productivity of life (Habitat, 2013; Van Nes, 2002). In spite of this, it has been suggested that, individual behavior change is also required to reduce private car use (Steg, 2005). Although people are willing to see problems solved, private car is still the priority choice because drivers think private car can provide their ideal service better (L. Wang et al., 2013). Luckily, a study shows that improving public transport service to meet requirement of passenger can lead to a large car use reduction (Eriksson et al., 2010). Moreover, improved public transport service can support to deliver sustainable city performance with objectives of economic efficiency and distributive justice (Hensher, 2008). Although heterogeneous mobility system provides higher possibilities and flexibilities of travelling, the provision of public transport is still constrained by inefficiencies in planning, operation, maintenance, which caused traffic- and service-related problems such as delay (Ramadan, 2016). Therefore, finding inefficiencies in heterogeneous mobility system and providing a competitive and efficient service is necessary to attract citizens to use public transport from private car.

However, finding inefficiencies in heterogeneous mobility system is complex and challenging due to the following three reasons. First, the multiple data sources are with different granularities, which is not easy to be aligned together and analyzed.

Second, inefficiencies in heterogeneous mobility system are due to multiple reasons from different aspects, which cannot be easily addressed by a single reason. Third and most importantly, inefficiency patterns are usually not well defined, which means automatic algorithms cannot detect unknown patterns with different variations.

Visualization is a technique to transform complex data to human-understandable graphical information. Combined with interactions, visualization can help users better understand the complex data features. Recently, visual analytics are proposed and fast developed, as a science of analytical reasoning facilitated by interactive visual interface (Thomas \& Cook, 2006). Spatial temporal visual analytics are techniques focusing on space and time related data (Andrienko et al. 2010). Visual analytics combines data mining techniques with visualization and interactions. It fully integrates human's capability for understanding complex scenarios and not well defined patterns. Considering the advantages of visual analytics and suitable scenarios, in this work, we propose to use visual analytics techniques to explore the inefficiency patterns in heterogeneous transportation system.

### 1.2 Research Objectives

Thanks to the availability of public transport data, including GPS data and schedule data, it is possible develop visual analytics methods to study and explore inefficiencies in mobility system.

Overall objective of this research is to discover factors contributing to potential inefficiencies in heterogeneous mobility system and build visual analytics methods to explore spatiotemporal patterns of those attributes. This objective can be achieved by following sub-objectives:

1. Finding factors contributing to inefficiencies of public transport service. Inefficiencies, result from individual or collective effect of factors affecting public transport service quality.
2. Building proper visual analytics methods to explore spatiotemporal inefficiency patterns. Each factor has its own characteristics. It is efficient to look into each factor and find proper method to visually explore the inefficiency patterns. In
addition, some factors may affect each other, it is also necessary to find relations between factors then process to visualization.
3. Implementing proposed methods with real data. In the case study, we use real data of Warsaw city to find inefficiencies patterns in Warsaw.

With the detailed summaries of inefficiencies with fine granularities and multiple aspects, we can better identify what are the interesting patterns of inefficiencies. Thus, we can design proper visual encodings and visual analytics techniques to address these inefficiencies. With the visualization, our target users, who are related to monitoring, operating, and analyzing transportation system, such as traffic specialist, city planners, can be better informed of the inefficiencies and understand the inefficiencies patterns. Thus, it can help them make data-driven decisions to improve the efficiency of the heterogeneous transportation system.

### 1.3 Structure of thesis

This paper is structured into 5 chapters. In chapter 1, we introduce the background of topic and objectives of the research. In chapter 2, we provide an overview of existing researches studying high quality public transport service and visual analytics of movement data. In chapter 3, we summarize possible factors contributing to inefficiencies of public transport from previous work and propose visualization methods to explore patterns of those factors. In chapter 4, we introduce the data used for the project, implement visualization methods with real data and show results as a case study. In chapter 5, we conclude the research methods and results, discuss limitation of study and future research.

## 2. State of the art

This chapter discusses previous work on public transport quality and visual analytics of movement data.

### 2.1 Key factors for efficient heterogeneous mobility system

With the growing demand of mobility and the diversity of travel needs, multiple modes of travelling have been introduced to transport system nowadays (Litman, 2011). Heterogeneous mobility system is the integration of multiple means of transportation (Beutel et al., 2014), including public and private use, such as trams, buses, taxis and ferries. In this thesis, heterogeneous mobility system refers to multimodal public transport for mass transit, including trams and buses.

Efficiency is the term used in measuring and evaluating the performance of organization (Mouzas, 2006). Efficiency of public transport systems can be analyzed based on factors relating to the quality of the service that is offered (Sampaio et al., 2008). Quality of service is also considered a transit efficiency indicators (Min et al., 2015)

The assessment and evaluation of public transport service quality can be measured from different perspectives. A research review (Redman et al., 2013) of quality attributes of public transport summarizes quality attributes that attract passengers changing from private car to public transport (Table 1). Attributes are described from two perspectives: physical and perceived. Physical attributes are measured without passengers while perceived attributes should be measured by observing passengers. In other word, physical attributes can be measured by data from vehicle operators such as GPS or timetable, while perceived attributes need to be measured by surveys on passengers.

|  | Attribute | Definition |
| :---: | :---: | :---: |
| Physical | Reliability | How closely the actual service matches the route timetable |
|  | Frequency | How often the service operates during a given period |
|  | Speed | How time spend travelling between specified points |
|  | Accessibility | The degree to which public transport is reasonable available to as many people as possible |
|  | Price | The monetary cost of travel |
|  | Information | How much information is provided about routes and |
|  | Provision | interchanges |
|  | Ease of | How simple transport connections of vehicles, including |
|  | Transfers or | time spent waiting |
|  | Interchanges |  |
|  | Vehicle | The physical and mechanical condition of vehicles, |
|  | Condition | including frequency of breakdowns |
| Perceive d | Comfort | How comfortable the journey is regarding access to seat, noise levels, driver handling, air condition |
|  | Safety | How safety from traffic accidents passengers feel during the journey as well as personal safety |
|  | Convenience | How simple the PT service is to use and how well it adds to one's ease of mobility |
|  | Aesthetics | Appeal of vehicles, stations and waiting areas to users' senses |

Table 1 Definitions of public transport service quality attributes (Redman et al., 2013)
The Transport Research Laboratory (Balcombe et al., 2004) categorizes attributes which can affect the service quality of public transport as attributes directly involving with time (access and alighting time, service intervals and in-vehicle time) and more problematical attributes (vehicle or rolling stock characteristics, interchange between modes, accessibility, information provision, marketing and promotion, and so on).

Murray (Murray, 2001) states that to increase the use of public transport and reduce private car, public transport service need to be with:

- more effective price structures;
- enhanced travel comfort;
- better suitability and convenience of service;
- reductions in travel time;
- increased service access.

This thesis focuses on physical attributes, excluding political, economical and technical attributes like price, information provision, speed and vehicle condition. We summarize other attributes affecting quality into time-related factors: reliability and frequency, and access-related factors: simplicity and accessibility.

### 2.1.1 Time-related factors

Travel time is an important consideration for passengers to choose a travel mode (Webb, 2011). Grdzelishvili \& Sathre (Grdzelishvili \& Sathre, 2011) reported that the main reason for car owners not using public transport is more efficient time use when travelling by car. Travel time consists of following components, which could be repeated at each transfer point:

- walking time to station
- waiting time
- in-vehicle time
- walking time to destination

These time components are perceived differently by passengers. Some researches (Van der Waard, 1988; Wardman, 2004) show that 1 minute waiting time is perceived as 1.5 minutes in-vehicle time by passengers. Compared with level of occupancy, access time, journey time, waiting time is the one that passengers are most willing to pay to reduce (Dell'Olio et al., 2017), which reveals waiting time is an important component in total travel time.

Under ideal situation, arrival and departure time of public transport should be fixed according to timetable. If passengers arrive in stations according to timetable, waiting time is manageable by themselves. However, problems in the real world such as congestions, fluctuation in demand and accidents lead to late arrival of vehicle or excessive in-vehicle time. It may lead to longer boarding time then increase the delay (Leong et al., 2016). Sometimes vehicles may arrive earlier than scheduled time so that people do not catch the vehicle in time and have to wait for next one for whole headway time (Ji et al., 2010).

The time difference between actual travel time for the trip and its official scheduled travel time can be defined as unreliability (Rietveld et al., 2001). Reliability has been considered as key attribute in determining public transport service quality (Redman et al., 2013). Cantwell et al. show that the lack of reliability induces stress because of long waiting time (Cantwell et al., 2009). Bhat \& Sardesai indicate that travel time reliability influence commuters travel mode decision because they attach high importance to the certainty of public transport (Bhat \& Sardesai, 2006).

There are many variations influencing the reliability of public transport. Bad weather may result in low speed of vehicles and congestion. Festival or events may attract the masses to a place and cause congestions. Accidents may stop vehicle moving and contribute to serious congestions. Consequently, the possibility of unreliability increases. But this kind of unreliability occur only at a certain time under certain conditions, which is the non-recurrent unreliability (Rietveld et al., 2001). Another unreliability called recurrent unreliability, which arises from fluctuations in demand and the manner in which transport is operated (Skabardonis et al., 2003). Nonrecurrent unreliability usually changes pattern of traffic all of a sudden (Yin \& Ieda, 2001), while recurrent unreliability shows more continuous patterns. This thesis focuses on recurrent unreliability.

Many researches have been done in order to find the cause of recurrent unreliability and the solution, such as behavioral response of drivers (Carey, 1998) or timetable constructions (Ceder, 1987). Some other researches focus on measurement of reliability. For example an estimation of unreliability of public transport chains in

Netherland (Rietveld et al., 2001), which gives an overall delay magnitude as well as probability and customers' valuation of unreliability.

There are few studies involving with spatiotemporal data of unreliability. Szymański et al. process large scale of vehicle positioning data and extract four profiles of stop to stop edges in public transport system to discover delay and delay change patterns (Szymański et al., 2017).

Another key factor influencing travel time is the frequency of service. Change of frequency may minimize the waiting time at the start of the trip as well as waiting time at transfer (Pratt et al., 2000). Many studies reveal frequency optimization particularly impacts the efficiency and demand of public transport. Ciaffi et al. proves that appropriate frequency setting can minimize fleet size under some constraints (Ciaffi et al., 2014). Increasing frequency has more influence on generating more journeys than reduced price (Brechan, 2017).

Impact of frequency changes with variations. In general, impact is higher when service frequency itself is lower. For example, the fact that frequency impacts on transit ridership change for buses has been found to be considerably greater on Sundays and in the evenings, when service frequency are generally lower, than other times (Preston, 1998). Similarly, elasticities tend to be higher in rural than in metropolitan areas, where service frequency are higher (Dargay \& Hanly, 1999). In off-peak hour, the impact of frequency is also higher than in the traffic hour. White et al. prove that when headway is long, passengers tend to arrive to station according to timetable. When headway is short, passengers tend to arrive in station randomly. In this case, average waiting time is the half of headway when vehicle is punctual, which indicates that the higher frequency is, the shorter waiting time is (White et al., 1992).

Because frequency impact on waiting time changes from different aspects with its own level, it makes sense to categorize frequency before studying it. Walker divides frequency in following three levels: The Frequent Network, which runs often enough that passengers do not need to plan travel by schedule (Walker, 2010). Headway is less than 15 minutes typically. Infrequent All-day services, which are the rest of the service that runs all day. Peak-only service, which exists only during the peak period.

In addition, regular interval also improves the efficiency of public transport because it is easier for passengers to remember and to facilitate transfer (Pucher \& Kurth, 1995).

In general, reliability and frequency are key factors affecting passengers' travel time. Unreliability can be recurrent or non-recurrent. Recurrent unreliability happens more regularly and arises from demand or operation of transport. Frequency impacts travel time with its own level, it makes sense to study frequency based on different frequency level.

### 2.1.2 Access-related factors

Public transport access is also focal service quality issue. Murray proves that service efficiency can be gained by improving extending service access and accessibility (Murray, 2003). In heterogeneous mobility system, when the coverage of public transport increases, access is not only the problem of the accessibility to a place from an origin but also how easy it is to get there.

Accessibility is the suitability for public transport taking individuals from origin to destination within a reasonable amount of time (Murray et al., 1998). Loader \& Stanley review the experience of improving accessibility in Melbourne by extending to outer areas with poor service level and providing weekend, evening service (Loader \& Stanley, 2009). This improved accessibility in both time and space contributes to the increase of bus ridership, which is historically high. In some case it even leads to users selling their cars.

In heterogeneous mobility system, transfer is necessary and inevitable especially when one access to rural area. Transfer allow lines or modes in heterogeneous mobility system to complement each other. In London, about 70\% underground trip and $30 \%$ bus trips include at least one transfer. Transfers enable passengers to travel into a multimodal network instead of a simple line. However, transfers cause inconvenience and complexity to passengers, and thus be less competitive than the car that goes from door to door (Z. Guo \& Wilson, 2011). A study (Bastidas-Zelaya \& Ruiz, 2016) in Quito found that one third of daily trips are using transfers, and the number of passengers decreases as the number of transfer increases.

### 2.2 Visual analytics of movement data

Traffic data has space time and attributes information, which makes the visualization and analysis of it profitable. Depending on different tasks, visual analytics of traffic data can be divided into: movement visualization over space, visualization of behavior along one trajectory over time, movement events extraction.

### 2.2.1 Movement visualization over space

In geographic space, many mobile objects move individually. Many researches are interested in getting information behind the large amount of movement, finding similarity or diversity of movement, comparing routes or movement efficiency, building overall view of movement over an area.

A work presented aggregated information of route diversity with given source/destination pairs over an area (Liu et al., 2011). It established some visual techniques to show not only time space but also attributes and statistics information from different views.

Zeng et al. provide three visualization techniques to analyze the mobility efficiency of passengers in public transportation system (Zeng et al., 2014). An isochrone map view shows the reachability from a selected start point within certain time. An isotime flow map view shows travel efficiency from a selected station to other stations within certain time. Upon the isotime flow map, an origin-destination (OD) pair journey view support to show mobility-related information along a selected route on a round-the-clock mobility wheel.

Density map, flow map, OD map are often used to visualize the aggregated information of movement. Willems et al. visualize a large number of vessel movement on a density map (Willems et al., 2009). Flow map are commonly employed to show the movement from one start position to different destinations (D. Guo, 2009). Flow maps can show large-volume data without causing visual clutter by merging edges (Phan et al., 2005).

### 2.2.2 Visualization of behavior along one trajectory over time

In real world, it is common that movement follow the same route over time. For example, in public transport system, one line goes back and forth through the same streets in the most of the time. However, some attributes along the route vary over time and across trips.

Sun et al. show traffic flow information of streets along time axis on a map. Traffic volume is displayed with graph on both sides of time axis and colors distinguish time and directions (Sun et al., 2017).

Space-time cube is often used to display limited number of trajectories. It is a3D cube with a geography reference on its base and time represented by its height. Typically it contains one space-time path with attributes shown by color (Hedley et al., 1999; Kraak, 2003).

### 2.2.3 Movement events extraction

By visualizing traffic data we can detect patterns. Those patterns can reveal events, such as incidents, congestions, which helps city planners to understand the context in which mobility happens.

Mazimpaka \& Timpf propose a method to discover and characterize significant locations and time periods along a bus routes effecting delay. It provides different views for understanding patterns and comparing across the space and time dimensions without clutter (Mazimpaka \& Timpf, 2016).
Z. Wang et al. present an interactive system for analyzing traffic jam. Traffic jam events are detected based on low-road-speed detection. A jam propagation graph, taking both spatially and temporally events into account, describes the emergence of a traffic jam and its influence over time (Z. Wang et al., 2013).

## 3. Methodology

In chapter 2, we studied existing work and found most existing work focuses on one part or aspect of efficiency in heterogeneous mobility system. In this thesis, we propose an overall feature summary to explore inefficiencies in heterogeneous mobility system from four aspects: unreliability, frequency, accessibility and complexity. For each aspect, we discover the feature of it and find proper visualization methods to help city planners understand inefficiencies better. The workflow of methodology is shown in Figure 1.


Figure 1 Workflow of methodology

### 3.1 Unreliability

As stated in chapter 2, reliability is one of the key attributes determining public transport service quality. Understanding where and when incurs unreliability and discovering patterns of it is very important for city planners to improve public transport performance. This problem is solved from three aspects: where and when
vehicle incurs delay along a route, where delay happens at a certain time, in general where and when delay happens most.

We define time difference between a vehicle's real arrival time in a station and scheduled arrival time in the same station as absolute difference, which reveals when and where has significant delay. But it would still be difficult for city planners to draw a conclusion about delay pattern. How delay changes in each place and time is also interesting. Therefore, we define time difference between absolute delay of two adjacent stations as relative difference, which can reveal how delay changes by time and position.

A bus or tram line runs back and forth along a route regularly over a period. At each stop along the route, there is an attribute difference. This attribute varies with time and position, but also shows patterns over a period. Those patterns may help city planners understand what happened to passengers in real context. For example, a line may always incur delay in peak off time at some stations. Then those hotspots are found close to big companies. Causes could be long boarding time and congestions caused by increasing number of private car on the street. Our goal is to detect those patterns of a certain line over a time.

To find those patterns, dynamic attributes along certain route as well as spatial and temporal information are needed to be presented. The challenge is to show that information at the same time. Main limitation of 2D solution is overlapping of trajectories because they have similar geometry. It would be difficult to explore individual trajectory and relation between trajectories. 3D visualization with time axis can avoid this problem. Our solution is based on (Tominski et al., 2012)'s stackingbased visualization of trajectory, trajectory wall.

Trajectory wall consists of a 2D map and stacking bands. Stacking bands are timeordered trajectories, which means the trajectory underneath starts earlier than the one above. Along one trajectory are colored segments, which show the attribute values by different color. Colors change of segments change along one trajectory, indicates the change of behavior along trajectory. 2D map serves as spatial reference underneath bands, offering spatial information of trajectory. An example showing speed of vehicle along the route can be seen in Figure 2. The average speed of vehicle is
$18 \mathrm{~km} / \mathrm{h}$. Speed where is close to average between $17 \mathrm{~km} / \mathrm{h}$ and $19 \mathrm{~km} / \mathrm{h}$, are shown in light yellow. Speed more than $19 \mathrm{~km} / \mathrm{h}$ are shown in red while lower than $17 \mathrm{~km} / \mathrm{h}$ are shown in blue. Arrows on trajectory indicate the direction of trajectory.


Figure 2 An example of trajectory wall showing speed of vehicle along the trajectory
Color selecting and interval setting are important because it conveys attribute information. First, we can categorize attribute of absolute time difference as:

- advance: absolute time difference is negative
- on time: absolute time difference is close to zero
- delay: absolute time difference is positive

For relative time difference, correspondingly, categories can be:

- decrease of delay: relative time difference is negative
- no influence on delay: relative time difference is close to zero
- increase of delay: relative time difference is positive

Then we set interval of each category. Values in different intervals will be later shown by different color to show the importance or degree of attribute. Two possibilities are proposed to set interval. One is based on perception of passengers, how much excessive time they are willing wait in general. For example, when passengers think
that delay within 5 minutes is tolerable but not more than 10 minutes, and vehicle should not departure more than 2 minutes in advance, 5 and 10 are set as breaks in delay category and -2 in advance category. The other possibility to set interval is based on headway. When delay time is close to headway, it is likely to incur bunching. Diverging color scheme, which emphasize data both above or below a critical value, is proper to discriminate individual categories and place equal emphasis (Brewer et al., 1997). Diverging color scheme has two different hues to distinguish negative value and positive value, light color in the middle to present values close to zero. Colors diverging from light to dark display degrees of delay, advance and increase or decrease of them. Four examples of diverging color scheme can be seen in Figure 3.


Figure 3 Diverging color scheme example (Brewer et al. 1997)
Transport situation changes by time and space. In weekdays, congestion may be more serious where close to working places during peak and peak off time, schools when students going to class. In weekend, congestion is more likely to happen close to shopping center or stadium. Delay is one of the symptoms of congestion (DoT, 2009). It can be thus reasonably assumed that delay shows different pattern by different time period (Pucha et al., 2007).

At any position, where public transport vehicle passes, there is an attribute value brought by vehicle. Over a period, like one hour, one day or one week, there is a series of attribute values. To describe the value of this position, probability-weighted average is applied with following steps:

1. Select time period of interest $\mathrm{T}_{\mathrm{m}}$, for example peak time and peak off time of a day, weekday and weekends of a week, summer and winter of a year.
2. Calculate excepted value of delay of each position along trajectories over selected time period. To simplify calculation, we approximate delay time into integer, for example, [ $0,0.5 \mathrm{~min}$ ) as 0 min , $[0.5 \mathrm{~min}, 1.5 \mathrm{~min}$ ) as 1 min . At position S , delay time $\mathrm{X}_{1}$ happens $\mathrm{N}_{1}$ times, $\mathrm{X}_{2}$ happens $\mathrm{N}_{2}$ times... $\mathrm{X}_{\mathrm{m}}$ happens $\mathrm{N}_{\mathrm{m}}$ times. The total number of delay records is N . Excepted value of delay $E_{k}$ at S during $\mathrm{T}_{\mathrm{m}}$ is:

$$
E_{k}=\sum_{m=0}^{N} X_{m} \frac{N_{m}}{N}
$$

3. Define influenced area of each stop.
4. Visualize excepted value of delay on map. One example showing speed attribute value can be seen in Figure 4. The excepted value of speed of tram 10 is encoded by color in polygons, which are areas where the vehicle runs through.


Figure 4 An example of visualization. Areas where tram 10 runs through are colored according to excepted value of speed when it passes. The $20^{\text {th }}, 40^{\text {th }}, 60^{\text {th }}, 80^{\text {th }}$ percentiles of the speed are colored by diverging scheme colors.

In public transport, individual transport mode may operate independently. Vehicles running along trails might have different traffic patterns with buses. We thus need to visualize and analyze each public transport mode separately.

Delay information of different transport mode over different periods are shown on different maps. It is convenient to explore delay pattern within certain time, but difficult to have an overall notion intuitively. For this reason, summarizing important information of all kinds of transport modes is necessary. From previous work we have few time periods of interest, into which any position can be classified. We use classification C to describe significant time period at one position. Time period of interest are set as $T_{1}, T_{2}$, whole time period as $T_{w}, E_{1}, E_{2}, E_{w}$ are respective excepted value of targeted attribute during $T_{1}, T_{2}$ and $T_{w}$ at position $S$. A threshold $E_{p}$ to
separate positions with high values from low ones. At each position S, we define with classification C as:

```
IF \(\quad \mathrm{E}_{\mathrm{w}}>=\mathrm{E}_{\mathrm{p}}\) THEN \(\mathrm{C}=\mathrm{T}_{\mathrm{w}}\)
ELSEIF \(\mathrm{E}_{1}>=\mathrm{E}_{\mathrm{p}}\) AND \(\mathrm{E}_{2}<\mathrm{E}_{\mathrm{p}}\) THEN \(\mathrm{C}=\mathrm{T}_{1}\)
ELSEIF \(\mathrm{E}_{2}>=\mathrm{Ep}_{\mathrm{p}}\) AND \(\mathrm{E}_{1}<\mathrm{E}_{\mathrm{p}}\) THEN \(\mathrm{C}=\mathrm{T}_{2}\)
ELSEIF \(\mathrm{E}_{1}>=\mathrm{E}_{\mathrm{p}}\) AND \(\mathrm{E}_{2}>=\mathrm{E}_{\mathrm{p}}\) THEN \(\mathrm{C}=\mathrm{T}_{1}+\mathrm{T}_{2}\)
ELSE C \(=\) NULL
```


### 3.2 Frequency

As mentioned in Chapter 2, frequency is another key factor that influence efficiency of public transport. In general, frequency setting should meet the demand of passengers and minimize passengers waiting time. As stated in Chapter2, when headway is short, passengers are assumed to arrive in stations randomly. In this case, waiting time is related to headway. When headway is long, passengers are intended to arrive in stations according to timetable, then waiting time would be minimized by improving reliability. Therefore frequency can be assessed from above three aspects: general situation, short headway and long headway. In general, we want to know where and when service frequency meets the demand of passengers or not. A threshold of demand frequency can be set to categorize frequencies. However, the demand of passengers varies by time and space. It is higher in day time than night time, higher in downtown than rural area. Therefore, we should discuss demand and threshold under space and time conditions.

When headway is short, which happens often in day time and downtown area, headway plays a determinative role in passenger waiting time. Wilson et al. (Wilson et al., 1992) prove that excepted passenger waiting time, when passenger arrival time is random, can be given by:

$$
\mathrm{w}=\frac{\mathrm{h}}{2}\left(1+c^{2}\right)
$$

Where
$\mathrm{w}=$ mean waiting time
$\mathrm{h}=$ mean headway
$\mathrm{c}=$ coefficient of variation, which is the ratio of the standard deviation to the mean headway

When headway is constant, waiting time is simply half of the headway. But when deviation of headway increases, so does the waiting time.

When headway is long, passengers tend to plan their travel according to timetable. Unreliability is the key factor that increase waiting time. By combing information of unreliability and low frequency, we can find where and when suffer from long waiting time most seriously.

### 3.3 Complexity and accessibility

A simple public transport mode is difficult to fulfill the travelling quality need of passengers, because the coverage of one simple mode is often small. Multi-modal public transport system, making individual mode supplement to each other, can often offer better accessibility to passengers. However, it is inevitable to make transfer in this case. Every transfer brings inconvenience like looking for station, which increases the complexity of travel. It can reasonably assume that passengers want most accessibility and least transfer times as possible.

Public transport system is often complex nowadays, for example in Warsaw, there are more than 6000 stations. It motivates us to visualize and analyze route from a common origin.

We use the combination of flow map and isochrone map to visualize. In a multi-trip travel, transfer is the connection between two trips, where passenger has to get off one vehicle and onto another. The number of transfer, or transfer times, is the number of connections in one travel. In Figure 5, the number of transfer of this travel is one. We
show different transfer times by qualitative colors and numbers of possible lines are shown by line thickness. Waiting time of each transfer point is shown in colored dot.

Accessibility can be discussed as space accessibility and time accessibility. In a multi model transport system, people can reach almost any station from a station by transferring. Hence, space accessibility from one station can be taken as the coverage of stations. Many studies apply 400 meters and 800 meters as service area around bus and tram(Furth \& Rahbee, 2000; Weinstein Agrawal et al., 2008). We thus set 400 meters around bus station and 800 meters around tram station are reachable areas. But it is not enough to assess accessibility without time because people will not choose public transport if it takes time more than expected. A common travel in a multimodel transport system is often the combination of few sub-travel (see Figure 3.2).


Figure 5 An example of travel with transfer
The time of the trip can be calculated as:
$\mathrm{T}_{\text {trip }}=\mathrm{T}_{\mathrm{w} 1}+\mathrm{T}_{\mathrm{t} 1}+\mathrm{T}_{\mathrm{t} 2}+\mathrm{T}_{\mathrm{w} 2}=\frac{\mathrm{L} 1}{\mathrm{Sw}}+\frac{\mathrm{L} 2}{\mathrm{St}}+\frac{\mathrm{L} 3}{\mathrm{Sb}}+\frac{\mathrm{L} 4}{\mathrm{Sw}}$

Tw1, Tt1, Tt2, Tw2 - travel time of walking one, trip one, trip two, walking two. L1, L2, L3, L4 —— distance of first walking one, trip one, trip two, walking two. Sw, $\mathrm{St}, \mathrm{Sb}$ —— speed of walking, tram and bus. Travel time range is defined as $(0$, $15 \mathrm{mins}]$, ( $15 \mathrm{mins}, 30 \mathrm{mins}$ ], ( $30 \mathrm{mins}, 45 \mathrm{mins}$ ]. Isochrone map shows area of equal travel time range from given origin by sequential colors. An example is presented in Figure 6.


Figure 6 Example of visualization of complexity and accessibility from given origin. Green circle shows waiting time at station, line shows times of transfer, blue area shows accessible area within certain time

## 4. Case study

In this chapter, proposed methods of unreliability and frequency are implemented by public transport data of city Warsaw. Inefficient time or places are discovered by visualization.

### 4.1 Data

In this section we introduced data and software used in case study and describe the detail information of data.

### 4.1.1 Data resource

The General Transit Feed Specification (GTFS) is a widely used common format for public transportation schedules and associated geographic information. GTFS feeds describes schedule, agency, route, stops data for fix-scheduled transport service in a series of related files. They allow public transit agencies to publish their transport data and developers to use published transport data in an interoperable way. TransitFeeds ${ }^{1}$ is a website which provides free service for developers to find open-source public transport data with GTFS format all around the world. For the purpose of this work, public transport data of Warsaw city on $1^{\text {st }}$ September was collected from TransitFeeds. We used python and V-analytics ${ }^{2}$ (Andrienko et al., 2007) software to generate timetable data.

GPS position points was provided by City of Warsaw within VaVeL project ${ }^{3}$. Further process was done by V-analytics software to generate real movement data, which includes trajectory forming, position id defining. Generated data description is shown in Table 2.

[^0]| Name | Example | Description |
| :--- | :--- | :--- |
| Id | $1 \_@ \_101 \_1 \_T O-$ <br> NWD_1 | Identify a segment along a trajectory. |
| Trajectory_id | $101 \_1 \_$TO-NWD_1 | Identify a trajectory |
| Line | 10 | Identify line |
| Tripmode | tram | Information of vehicle, bus or tram |
| Start_ID | 1462 | Identify start point of a segment |
| End_ID | 1122 | Identify end point of a segment |
| Start_time | $2016090100: 00: 00$ | Starting time of the segment |
| End_time | $2016090100: 02: 00$ | Ending time of the segment |
| Segment N | 1 | The sequence number of the segment along the <br> trajectory |
| Start_X | 20.96036911 | Longitude of start point |
| Start_Y | 52.31063843 | Latitude of start point |
| End_X | 20.96449089 | Longitude of end point |
| End_Y | 52.3132515 | Latitude if end point |

Table 2 Description of data
There are more than 6000 stops in Warsaw. Many of stops are very close. Some stops are the same stop with different directions. Sometimes one station is distributed closely intended to divide vehicles or traffic flows. It would save a lot of work to cluster them and represent by one stop. Start_ID or End_ID is an id representing a group of stops which are divided by voronoi polygons (Figure 7), which shows spatial close of points within it.


Figure 7 Transport stops in Warsaw city. Blue points are original stops in Warsaw and red points are centroids of a group of points inside an voronoi polygon.

### 4.1.2 Data processing

Our data contains arrival time information about when vehicle should arrive in station in timetable data and when it actually arrives in real movement data. A series of arrival time differences between points along timetable trajectories and real movement trajectories can describe the unreliability of the line, this is also one of the goals of the data processing. We can perform data processing as two steps: first: matching corresponding trajectories, second: calculate the time difference between points along trajectories.

Matching corresponding trajectories can be interpreted into finding corresponding trips. Ideally, each trajectory from timetable should correspond with a real GPS trajectory and each trajectory is a trip. We can put trips from timetable and real
movement in sequence of time, and then match them one by one. However, in real move data, one trajectory often contains few trips (Figure 8).


Figure 8 Comparison between timetable data and real movement data. Trajectories of tram 10 on 01.09.2016 are shown in space time cube, timetable ones are in orange and real movement ones are in blue. One selected trajectory is shown in black.

To segment trajectory, many criteria have been studied and proposed, such as speed, curviness and sharp angle (Buchin et al., 2011). However, there are many lines with different geographic and temporal features, which makes it difficult to set common criteria. In addition, any mistake may cause mismatching rest of trips. Therefore, it is necessary to use other characteristic of data to match trajectories.

Corresponding trajectories in timetable and real move should have geographical and temporal similarity. Geographically, it should follow the same or similar routes, which in this case is the same sequence of Start_ID and End_ID. Temporally, the start time of each trip and duration between stations should be very close. We propose two methods to match corresponding trajectories, one is point-based, the other is trajectory-based.

A line running along a fixed route always goes through certain positions. One position has a sequence of time records when a line goes through. At each position, we can match time record of timetable and real move one by one.

In space time cube, we can describe a trajectory route, starting time, speed, direction. Because trajectories of one traffic line often follow the same route, we only need to take starting time, speed, and direction into account, which are already in the data.

We also notice that there are some problems with real move data, jump positions and redundant records. Jump position is the position record of a station that is far away from the real station, which is corrected by python. Redundant record is more records between two real stations. Those redundant records are deleted by python.

The result of trajectory matching can be seen in Figure 9. The comparison with enlarged area, which is marked in red square in Figure can be seen in Figure 10. Point-based result has jump position and wrong time sequence, it happens that at one time there are two positions in one trajectory. Trajectory-based result keeps original geographic feature and time sequence better. We use trajectory-based result for later work.


Figure 9 Comparison between trajectory matching results. a1. original tram 10 trajectories of timetable. b1. point-based result of tram 10 b 2 . c. trajectory-based result of tram 10 . a2 b2 c2. Corresponding trajectories shown in space time cube and black line is one selected trajectory


Figure 10 Enlarged comparison between trajectory matching results. a. original tram 10 trajectories of timetable. b. point-based result of tram 10. c. trajectory-based result of tram 10. Black line is one selected trajectory. In point-based result, it happens that at one time there are more than one position in a trajectory, which are marked in red square.

### 4.2 Visualization of unreliability

The software used to visualize the data was V-Analytics ${ }^{4}$ (Andrienko et al., 2007).
We selected tram line 10 as an example to visualize unreliability along a route because it is one of the main lines which through city center. The data of tram 10 has 173 trajectories in total. The number of stops of tram 10 according to timetable is between 14 and 39 , the duration is between 25 minutes to 55 minutes. Considering of some stops may lose in real movement data, trajectories with less than 10 stops and less than 20 minutes duration, which are assumed to be incomplete, are filter out. About 23\% data was filtered and 133 trajectories are left to visualize.

Absolute time difference of tram 10 ranges from -551 seconds to 726 seconds. It is assumed that, time difference within 1.5 minutes ( 90 seconds) can be ignored, 5 minutes is the limitation of tolerable waiting time due to late arrival. Based on this assumption, we set interval as (-×, -300], (-300, -90], (-90, 90], (90, 300], (300, + - ).

[^1]We use red color to represent delay, light yellow to represent on time and green to represent advance, which are most commonly recognized colors that show traffic conditions. We visualize the data of absolute time difference in a trajectory wall (Figure 11).


Figure 11 Trajectory wall of absolute time difference of tram 10
The trajectory wall presents tram 10 behavior from 04:14 to 23:59 on $1^{\text {st }}$ September of 2016 in Warsaw. Trip started at $04: 14$ shows on the bottom of temporally stacked trajectories while trip ended at 23:59 shows on the top. The real position of each point along trajectory can be obtained by projecting trajectory on the map. We observe that delay distribution varies over time but more densely during certain times. By hovering the trajectory of interest, we can get time information on the left side of trajectory wall. We note that during 08:30-09:30 and 17:00-18:00 delay happens more frequently (Figure 12). Furthermore, advance distribution is more stable over the day. We can find the position where advance incurs often by shifting perspective (Figure 13). Then we visualize relative time difference to discover where increase and decrease delay.

a

b

Figure 12 Trajectory of interest is stroked by white lines. Significant delay is marked in blue square. Temporal information can be seen at the left side. $a$ is at 08:58. $b$ is at 17:34


Figure 13 Perspective-shifted trajectory wall. By shifting perspective we can see position on the real map

Then we visualized relative time difference to discover where increase and decrease delay (Figure 14). We set interval at $-180,-60,60,180$ seconds, and use red to represent delay increase while green to represent delay decrease. It can be seen that delay and advance show spatial feature. By zooming the trajectory wall we note that feature varies with direction (Figure 15). We thus selected and visualized trajectories by directions. The result shows clear spatial patterns of individual direction (Figure 16). An example of pattern is marked in blue square, where delay increase or decrease happens repetitively over a period.


Figure 14 Trajectory wall of relative time difference


Figure 15 Enlarged trajectory wall of relative time difference. Grey arrows on trajectory band indicate the direction of trajectory. In blue marked square, the same place within close time range, different directions show opposite relative time difference, which red is positive while green is negative.

a


Figure 16 Trajectory wall with directions. a. Trajectory wall of tram 10 of direction to OsGczewska. b. Trajectory wall of tram 10 of direnction to Wycigi. In blue marked square, delay increase or decrease happens repetitively over a period.

We notice that delay and delay increase pattern changes by time, but in a period, it keeps some features. Figure 12 presents during time 8:30-9:30 and 17:00-18:00 delay happens more frequently. We assume these two periods as peak time. To find delay pattern in peak time, we extracted trajectories during peak time, and compared them with the average delay of the day. Different transport modes have different reactions
to traffic situation and problems. Here we compare delay and delay increase of tram and bus over the same time period (Figure 17, Figure 18).

In Figure 17, comparing a, c, e, it is notable that delay is more serious in North-South direction in a and c , which means in peak time trams with direction from north to south or south to north may have more serious problem with delay. Comparing b, d, f, it is notable that delay happens more in areas outside center in d while more in center areas in f , which means bus may have more delays in suburban in morning peak time while more in center areas in evening peak time.

In Figure 18, there are some significant places where increase much delay, which are colored in red. It also worth noting that there are some places, where delay increase all the time, which are marked in blue.

As mentioned in Chapter 2.1, we calculated expected value of delay and delay increase of each stop id. Each stop id is associated with a voronoi polygon. In other words, one voronoi polygon is area of influence of the corresponding stop id. We thus visualize the attribute value of the stop in the associated influencing area.

After comparison, we then summarize significant places of delay and delay increase. As stated in chapter 3.1, we set a threshold to separate significant positions from all positions, which is 3 minutes for delay and 2 minutes for delay increase. At one position, has expected value of delay $\mathrm{E}_{1}$ during 8:30-9:30, $\mathrm{E}_{2}$ during 17:00-18:00, $\mathrm{E}_{\mathrm{w}}$ for the whole day, the classification of each transport mode has five possible values:

```
IF E Ew >= 3min THEN C = All day
ELSEIF E > >= 3min AND E2 < 3min THEN C = 8:30-9:30
ELSEIF E > >= 3min AND E1 < 3min THEN C = 17:00-18:00
ELSEIF E > >= 3min AND E2 >= 3min THEN C = peak time
ELSE C = NULL
```

The interesting finding in the result is that all the classification of peak time is already included in all day class, which means significant delay places in peak time is also show significant delay when time scope extends to a whole day.

Therefore, four classifications are assigned to one transport mode at one position. We discuss bus and tram modes. Combing both modes, except null and null combination, we have 15 possible classes in all. The same process was also implemented on delay increase data. The result can be seen in Figure 19. Red color represents tram and green color represents bus, brown color, which is mix of green and red, represent both tram and bus. In Figure 19.a, we notice that red colors show more in north and south in the city, while green show more in suburban, brown show more in center. This result is compatible with the result shown in Figure 17. In Figure 19.b, significant delay increase of different time and transport modes can be seen.


Figure 17 Delay comparison. Delay of a. tram during 8:30-9:30 b. bus during 8:30-9:30 c. tram during 17:00-18:00 d. bus during 17:00-18:00 e. tram of the whole day f. bus of the whole day. Legend in a applies to also $\mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}$, and f


Figure 18 Delay increase comparison. Delay increase of a. tram during 8:30-9:30 b. bus during 8:30-9:30 c. tram during 17:00-18:00 d. bus during 17:00-18:00 e. tram of the whole day f . bus of the whole day. Legend in a applies to also $\mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}$ and f


Figure 19 Summary of delay and delay increase. a. summary result of delay b. summary result of delay increase.

### 4.3 Visualization of frequency

Frequency setting should meet the demand of passengers. As the demand varies by time and position, the threshold of frequency setting should be dependent on demand dynamically. We take day time, night time, route running though city center or not as variables.

It is obvious that demand of transport is higher in day time than night time, higher with routes though city center than bypass city center. We thus set different assumed threshold under different conditions.

In Figure 20 frequency during 07:00-19:00 on 2016-09-01 is presented. We set threshold at 15 minutes headway for routes thought city center, which is marked in blue border, 30 minutes headway for other routes. Routes with headway shorter than threshold, which means meet the demand of passengers, are colored in light yellow. Routes with headway longer than threshold, which means not meet the demand, are shown in red. Tram lines are shown in solid lines while bus lines are in dash lines.

In Figure 21 frequency during 22:00 on 2016-09-01 to 5:00 next day is presented. We set threshold at 30 minutes headway for routes thought city center, which is marked in blue border, 60 minutes headway for other routes. Routes with headway shorter than threshold are colored in light yellow while routes with headway longer than threshold are shown in red. Tram lines are shown in solid lines while bus lines are in dash lines.


Figure 20 Frequency during 07:00-19:00 on 2016-09-01


Figure 21 Frequency during 22:00 on 2016-09-01 to 5:00 next day

It can be seen in Figure 20 and Figure 21 that, routes close to city center, has higher frequency than routes in rural area. In other word, public transport close to city center has higher possibility to meet passengers demand.

Then we explored high frequency data to find mean waiting time. All the trams with headway less than 15 minutes during 13:30-14:30 time period were selected, as it has less delay problem at this time from previous result. Then mean waiting time was calculated by method proposed in Chapter 3.2. The result is shown in Figure 22.


Figure 22 Mean waiting time of high frequency trams during 13:30-14:30
When frequency is low, reliability of transport is determinative in reducing waiting time. We visualize unreliability and low frequency at the same time to find significant places, where have serious delay and frequency of transport is low. From previous work we know that during 8:30-9:30, delay is relatively serious. Tram routes with headway shorter than 15 mins and bus routes shouter than 30 minutes are filtered. Threshold for buses and trams are different because more buses usually run away
from city and have longer headway. Delay degree is shown by color and frequency shown by line thickness. The result can be seen in Figure 23 and Figure 24. Significant places where frequency is low and unreliability is high is marked in red square.


Figure 23 Combination of tram frequency and delay


Figure 24 Combination of bus frequency and delay

## 5. Discussion

After implementing the methodology in the case study, we discuss the pros and cons of our method in the scope of understanding inefficiencies with visualization in heterogeneous mobility system.

Our case study generally confirms the capability of our visual analytics approach to find inefficiency patterns of unreliability and frequency. In the aspect of unreliability, our method finds such patterns in terms of space and time. First, we find delay peak time in both morning 08:30-09:30 and evening 17:00-18:00 in the workday. This makes sense to relating such patterns with the general rushing hours. Secondly, we can visually identify the delays in terms of spatial context. We find that trams experience more delay in north and south in city while buses experience more in east and north. Moreover, buses show more delay patterns in suburban areas in the morning and more delay patterns in the city center in the evening. Visualizations can help users quickly identify the peak time and regions with interesting patterns. Thus, these regions should be focused to improve the inefficiencies for city planners.

In the aspect of frequency, we allow users to set different filtering parameters to identify the demand of a specific transportation mode in different space and time. In the case study, we set two parameters. One is the traffic time, and the other is whether the vehicles are passing the inner city. Thus, with the visualization of different filtering parameters, the frequency of specific traffic mode in specific space and time can be identified. For example, we highlight the trajectories passing inner cities and not in the morning. Users can find these trajectories which are less frequent than the threshold values, which indicate the transportations in the regions and time should be focused and improved. With such methods, users such as urban planners can add more public transportation lines within the identified regions. Moreover, we allow users to identify the mean waiting time of traffic lines with high frequency. It further enables the targeting users to improve the efficiency according to the identified regions with long waiting time.

One step further, we conduct correlated analysis and visualization by combining frequency and unreliability. With this approach, places where have low frequency and high unreliability were discovered. Those places need to pay more attention because it indicates high inefficiencies in public transport system.

Our approach has disadvantages. In this stage, we didn't implement the case study of complexity and accessibility due to the restriction of data availability. We theoretically identify the different modes with visualization design and argue that with visualization, multiple modes of transportation can be fused and visualized on the same map. It helps users to grasp the features with low accessibility and high complexity from one origin on the map.

In the case study, we only consider two transportation modes, which are trams and buses. We envision to extend it to multiple transportation modes with higher scalability. Furthermore, there are still occlusions in the visualization, we need to improve it in the future.

## 6. Conclusion

Finding inefficiencies in heterogeneous mobility system is needed to attract passengers to use public transport. Previous works mostly focus on one aspect of inefficiency. This thesis aims at understanding overall features of inefficiencies in heterogeneous mobility system by visual analytics methods. First, we summarize four attributes affecting public transport inefficiencies based on previous works: unreliability, frequency, complexity and accessibility. Then we propose visual analytics method for individual attribute. For unreliability, we explore it from trajectory level, regional level and found peak time distribution. In trajectory level, trajectory wall is used to show spatial temporal information, from which we find delay peak time of weekday. In regional level, expected value of delay is calculated to show the probability of delay at certain position. Afterwards we visualize expected value of delay over an area to compare delay probability of different positions. In the case study, we find that trams and buses show different delay patterns in different locations and time. By implementing peak time distribution, significant delay time of a certain position is discovered.

For frequency, we allow users to filter the frequency of specific transport mode where does not meet the demand. When frequency is high, we provide possibilities to discover regions where waiting time of transport is high. A visualization of correlated attributes is done to find significant inefficient positions with high unreliability and low frequency.

For complexity and accessibility, a visualization method consists of transfer times, spatial and temporal accessibility was proposed. However, due to restrictions of data, this method is not implemented in the case study.

In summary, we address on the inefficiency issues in heterogeneous mobility system. With the detailed summary and analysis on four aspects of inefficiency issues, we propose visual analytics approaches to enable targeting users to better understand the inefficiency issues of the heterogeneous mobility system. The visualization reflects multiple aspects of features according to space and time. Users, such as city planners,
can quickly realize which location and with which period of time the city experience the inefficiency. Furthermore, with the understanding of each features, e.g. unreliability, frequency, complexity and accessibility, the users are expected to propose methods to solve the inefficiency issues in specific regions, time and transportation modes. With such human-in-the-loop visualization approach, our case study confirms the capability of proposed method for visualizing and understanding the inefficiencies.

## Reference

Andrienko, G., Andrienko, N., \& Wrobel, S. (2007). Visual analytics tools for analysis of movement data. ACM SIGKDD Explorations Newsletter, 9(2), 38-46.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., . . . White, P. (2004). The demand for public transport: a practical guide.

Bastidas-Zelaya, E., \& Ruiz, T. (2016). Analysis of Multistage Chains in Public Transport: The Case of Quito, Ecuador. Transportation Research Procedia, 18, 180188.

Beutel, M. C., Samsel, C., Mensing, M., \& Krempels, K.-H. (2014). Business model framework to provide heterogeneous mobility services on virtual markets. Paper presented at the e-Business (ICE-B), 2014 11th International Conference on.

Bhat, C. R., \& Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological, 40(9), 709-730.

Brechan, I. (2017). Effect of Price Reduction and Increased Service Frequency on Public Transport Travel. Journal of Public Transportation, 20(1), 8.

Brewer, C. A., MacEachren, A. M., Pickle, L. W., \& Herrmann, D. (1997). Mapping mortality: Evaluating color schemes for choropleth maps. Annals of the Association of American Geographers, 87(3), 411-438.

Buchin, M., Driemel, A., Van Kreveld, M., \& Sacristán, V. (2011). Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. Journal of Spatial Information Science, 2011(3), 33-63.

Cantwell, M., Caulfield, B., \& O'Mahony, M. (2009). Examining the factors that impact public transport commuting satisfaction. Journal of Public Transportation, 12(2), 1.

Carey, M. (1998). Optimizing scheduled times, allowing for behavioural response. Transportation Research Part B: Methodological, 32(5), 329-342.

Ceder, A. (1987). Methods for creating bus timetables. Transportation Research Part A: General, 21(1), 59-83.

Ciaffi, F., Cipriani, E., Petrelli, M., \& Uspalyte-Vitkuniene, R. (2014). A new methodology for the public transport network design. Paper presented at the Environmental Engineering. Proceedings of the International Conference on Environmental Engineering. ICEE.

Dargay, J., \& Hanly, M. (1999). Bus fare elasticities. Report to the UK Department of the Environment, Transport and the Regions. London, ESRC Transport Studies Unit, University College London, 132.

Dell'Olio, L., Ibeas, A., Oña, J. d., \& Oña, R. d. (2017). Chapter 10 - Beyond Perceived Quality: Desired Quality. In Public Transportation Quality of Service (pp. 181-194): Elsevier.

DoT, U. (2009). Assessing the Full Costs of Congestion on Surface Transportation Systems and Reducing Them through Pricing. United States Department of Transportation.

Eriksson, L., Nordlund, A. M., \& Garvill, J. (2010). Expected car use reduction in response to structural travel demand management measures. Transportation research part F: traffic psychology and behaviour, 13(5), 329-342.

Furth, P., \& Rahbee, A. (2000). Optimal bus stop spacing through dynamic programming and geographic modeling. Transportation Research Record: Journal of the Transportation Research Board(1731), 15-22.

Goodwin, P. (1996). Simple arithmetic. In: Pergamon.
Grdzelishvili, I., \& Sathre, R. (2011). Understanding the urban travel attitudes and behavior of Tbilisi residents. Transport policy, 18(1), 38-45.

Greene, D. L., \& Wegener, M. (1997). Sustainable transport. Journal of Transport Geography, 5(3), 177-190.

Guo, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. IEEE transactions on visualization and computer graphics, 15(6).

Guo, Z., \& Wilson, N. H. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. Transportation Research Part A: Policy and Practice, 45(2), 91-104.

Habitat, U. (2013). State of the world's cities 2012/2013: Prosperity of cities: Routledge.

Hedley, N. R., Drew, C. H., Arfin, E. A., \& Lee, A. (1999). Hagerstrand revisited: Interactive space-time visualizations of complex spatial data. INFORMATICA-LJUBLJANA-, 23, 155-168.

Hensher, D. (2008). Frequency and connectivity: key drivers of reform in urban public transport provision. Journeys, Nov.

Ji, Y., Mishalani, R. G., \& McCord, M. R. (2010). Analytical and empirical investigations of the effect of bus drivers' reactions to schedules on transit operations reliability. Paper presented at the Proceedings of the 12th world conference on transportation research, Lisbon.

Kraak, M.-J. (2003). The space-time cube revisited from a geovisualization perspective. Paper presented at the Proc. 21st International Cartographic Conference.

Leong, W., Goh, K., Hess, S., \& Murphy, P. (2016). Improving bus service reliability: The Singapore experience. Research in Transportation Economics, 59, 40-49.

Litman, T. (2011). Introduction to multi-modal transportation planning. Victoria Transport Policy Institute, 15.

Liu, H., Gao, Y., Lu, L., Liu, S., Qu, H., \& Ni, L. M. (2011). Visual analysis of route diversity. Paper presented at the Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on.

Loader, C., \& Stanley, J. (2009). Growing bus patronage and addressing transport disadvantage-The Melbourne experience. Transport policy, 16(3), 106-114.

Mazimpaka, J. D., \& Timpf, S. (2016). A visual and computational analysis approach for exploring significant locations and time periods along a bus route. Paper presented at the Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science.

Min, H., Ahn, Y.-H., \& Lambert, T. E. (2015). Evaluating the comparative efficiency of urban mass transit systems: a longitudinal analysis of the Ohio case. International Journal of Logistics Research and Applications, 18(6), 518-534.

Mouzas, S. (2006). Efficiency versus effectiveness in business networks. Journal of Business Research, 59(10), 1124-1132.

Murray, A. T. (2001). Strategic analysis of public transport coverage. Socio-Economic Planning Sciences, 35(3), 175-188.

Murray, A. T. (2003). A coverage model for improving public transit system accessibility and expanding access. Annals of Operations Research, 123(1), 143-156.

Murray, A. T., Davis, R., Stimson, R. J., \& Ferreira, L. (1998). Public transportation access. Transportation Research Part D: Transport and Environment, 3(5), 319-328.

Phan, D., Xiao, L., Yeh, R., \& Hanrahan, P. (2005). Flow map layout. Paper presented at the Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on.

Pratt, R. H., Turnbull, K. F., Evans, I., John, E., McCollom, B. E., Spielberg, F., . . . Kuzmyak, J. R. (2000). Traveler response to transportation system changes: Interim handbook. Retrieved from

Preston, J. (1998). Public transport elasticities: time for a re-think? TRANSPORT STUDIES UNIT OXFORD UNIVERSITY-PUBLICATIONS-TSU-ALL SERIES-.

Pucha, H., Zhang, Y., Mao, Z. M., \& Hu, Y. C. (2007). Understanding network delay changes caused by routing events. Paper presented at the ACM SIGMETRICS Performance Evaluation Review.

Pucher, J., \& Kurth, S. (1995). Verkehrsverbund: the success of regional public transport in Germany, Austria and Switzerland. Transport policy, 2(4), 279-291.

Ramadan, A. (2016). INFORMAL TRANSPORT IN MENA, WE DON'T KNOW ENOUGH. MENA CTE Journal, 35-41.

Redman, L., Friman, M., Gärling, T., \& Hartig, T. (2013). Quality attributes of public transport that attract car users: A research review. Transport policy, 25, 119-127.

Rietveld, P., Bruinsma, F. R., \& Van Vuuren, D. J. (2001). Coping with unreliability in public transport chains: A case study for Netherlands. Transportation Research Part A: Policy and Practice, 35(6), 539-559.

Sampaio, B. R., Neto, O. L., \& Sampaio, Y. (2008). Efficiency analysis of public transport systems: Lessons for institutional planning. Transportation Research Part A: Policy and Practice, 42(3), 445-454.

Skabardonis, A., Varaiya, P., \& Petty, K. (2003). Measuring recurrent and nonrecurrent traffic congestion. Transportation Research Record: Journal of the Transportation Research Board(1856), 118-124.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. Transportation Research Part A: Policy and Practice, 39(2), 147-162.

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.

Szymański, P., Żołnieruk, M., Oleszczyk, P., Gisterek, I., \& Kajdanowicz, T. (2017). Spatio-temporal profiling of public transport delays based on large scale vehicle positioning data from GPS in Wroc $\{\backslash 1\}$ aw. arXiv preprint arXiv:1707.07913.

Thomas, J. J., \& Cook, K. A. (2006). A visual analytics agenda. IEEE computer graphics and applications, 26(1), 10-13.

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.

Van der Waard, J. (1988). The relative importance of public transport trip time attributes in route choice. Paper presented at the PTRC Summer Annual Meeting, 16th, 1988, Bath, United Kingdom.

Van Nes, R. (2002). Design of multimodal transport networks: A hierarchical approach.

Walker, J. (2010). [Web log post]. Retrieved from http://humantransit.org/2010/08/basics-the-case-for-frequency-mapping.html

Wang, L., Li, L., Wu, B., \& Bai, Y. (2013). Private Car Switched to Public Transit by Commuters, in Shanghai, China. Procedia-Social and Behavioral Sciences, 96, 12931303.

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.

Wardman, M. (2004). Public transport values of time. Transport policy, 11(4), 363377.

Webb, R. (2011). Public transport versus the private car: a study of attitudes towards transport modes in Plymouth. The Plymouth Student Scientist, 4(1), 325-371.

Weinstein Agrawal, A., Schlossberg, M., \& Irvin, K. (2008). How far, by which route and why? A spatial analysis of pedestrian preference. Journal of urban design, 13(1), 81-98.

White, P. R., Turner, R., \& Mbara, T. (1992). Cost benefit analysis of urban minibus operations. Transportation, 19(1), 59-74.

Willems, N., Van De Wetering, H., \& Van Wijk, J. J. (2009). Visualization of vessel movements. Paper presented at the Computer Graphics Forum.

Wilson, N. H., Nelson, D., Palmere, A., Grayson, T. H., \& Cederquist, C. (1992). Service-quality monitoring for high-frequency transit lines. Transportation Research $\operatorname{Record}(1349)$.

Yin, Y., \& Ieda, H. (2001). Assessing performance reliability of road networks under nonrecurrent congestion. Transportation Research Record: Journal of the Transportation Research Board(1771), 148-155.

Zeng, W., Fu, C.-W., Arisona, S. M., Erath, A., \& Qu, H. (2014). Visualizing mobility of public transportation system. IEEE transactions on visualization and computer graphics, 20(12), 1833-1842.


[^0]:    ${ }^{1}$ http://transitfeeds.com/
    2 http://geoanalytics.net/V-Analytics/
    ${ }^{3}$ http://www.vavel-project.eu/

[^1]:    ${ }^{4}$ http://geoanalytics.net/V-Analytics/

