Department of Computer Science
University of Houston

A Visual Analytic System for Longitudinal Transportation Data of Great Britain

Marzieh Berenjkoub*, Harsha Nyshadham*, Zhigang Deng*, and Guoning Chen*

Department of Computer Science
University of Houston
Houston, TX, 77204, USA
http://www.cs.uh.edu

Technical Report Number UH-CS-15-01

July 9, 2015

Keywords: Public transportation data, visual analytics.

Abstract

It is a known fact today, that globalization has increased the opportunities around the world and more so in the urban areas. As a result, there is high concentration of population in all urban communities around the world. In order to meet these needs, public transportation systems have been developed at a fascinating pace around the world. This leads to a vast amount of transportation data. Now researchers and practitioners are digging in to find new ways to improve it further. In this paper, we explore the longitudinal transportation data of Great Britain and visually compare changes and trends over the years. We try to answer to what extent the transportation network has directly impacted the population and job distribution in major cities of Great Britain, and vice versa. We also aim to evaluate potential issues in the current services, such as the potential accumulated delay due to the overlapping services. We provide flexible search queries, allowing users to interact with different aspects of city geography and dynamically see the changes taking place at those locations. Finally, we present a series of case studies and report some interesting findings.

*Marzieh Berenjkoub, Harsha Nyshadham, Zhigang Deng, and Guoning Chen are with University of Houston

A Visual Analytic System for Longitudinal Transportation Data of Great Britain

Marzieh Berenjkoub*, Harsha Nyshadham*, Zhigang Deng*, and Guoning Chen*

Abstract

It is a known fact today, that globalization has increased the opportunities around the world and more so in the urban areas. As a result, there is high concentration of population in all urban communities around the world. In order to meet these needs, public transportation systems have been developed at a fascinating pace around the world. This leads to a vast amount of transportation data. Now researchers and practitioners are digging in to find new ways to improve it further. In this paper, we explore the longitudinal transportation data of Great Britain and visually compare changes and trends over the years. We try to answer to what extent the transportation network has directly impacted the population and job distribution in major cities of Great Britain, and vice versa. We also aim to evaluate potential issues in the current services, such as the potential accumulated delay due to the overlapping services. We provide flexible search queries, allowing users to interact with different aspects of city geography and dynamically see the changes taking place at those locations. Finally, we present a series of case studies and report some interesting findings.

Index Terms

Public transportation data, visual analytics.

I. Introduction

Public transportation is increasingly becoming the preferred way to travel and commute for people around the world. Globalization has increased the population in almost every urban area. In most cities population density tends to grow towards the outer suburbs [29], making public transportation more essential. While globalization has increased the population in almost every urban area, traffic congestion has become a major issue in our communities today. People have started to prefer using public transport to commute between jobs and their homes. Keeping up with this, new services are being added every day to enable people to have more convenient and faster travel facilities. The transportation system of a city tells the general sense of lifestyle in the city today.

Recently, such data have been made available to the general public [20], [23]. These data include geospatial and temporal attributes, and with the visualization resources available today, they give us a lot of opportunities to explore these spatio-temporal data and to come up with interesting findings that are difficult to obtain otherwise [6], [9]. Visualizing longitudinal data of this kind is challenging because of the multi-dimensional and tightly linked data [27]. For example, a route is comprised of a couple of journeys a service takes, where each journey passes through many stops. In the same way, all these stops are a part of many routes. Therefore, the public networked transportation data are more complicated than trajectory data [8], [10], [30]. In addition, the time schedule of such data has a large number of overlaps, which gives scope for wide visualization and analysis tasks.

In this paper, we aim to help analysts to understand the correlation between public lifestyle and a city's transportation service. In particular, we focus on the transportation system of Great Britain [3]. The important questions we seek answers for include: a) is an increase in the transportation network a result of an increase in jobs or population? b) If yes, to what extend does transportation system contributes to this trend? c) can we summarize the change of the lifestyle of the residents of a city based on the trend of the transportation services over the years? d) will the increase of transportation services lead to potential traffic congestion and where are those hot spots? Existing systems [17], [21] that are solely focused on the analysis of the transportation data are not sufficient to answer these questions.

We address these questions by designing a visual analytic system to visualize the transportation data as well as the population and job density of the Great Britain from 2004 to 2011. Our system helps users look at different aspects of a transportation network. Users can look for transportation connectivity of a single location, to group of

*Marzieh Berenjkoub, Harsha Nyshadham, Zhigang Deng, and Guoning Chen are with University of Houston

locations, to a region of the city and compare it with past years. Users can quickly correlate this with population and job distribution and see the trend over time. Our tool can also be used by domain experts who are involved in the designing and planning of transportation system. It helps them analyze whether the services provided at any stops are reasonable or not, and supports the study of possible traffic congestion at certain time and possible time overlap between different services.

II. Related Work

During the past decade a lot of research was focused on transportation networks. Many applications, such as traffic simulation [12], road traffic surveillance system [14], and travel planners [2], [18], [28], were developed to help domain experts and government agencies understand the structure of transportation or help commuters make plans given live traffic information. Various traffic data are generated from these applications. Visualization is one effective tool for the interpretation of these data.

Visualization techniques: One of the key challenges in working with traffic data is its temporal attributes. Aigner et al. [5] have considered time as a new dimension; they used principal component analysis for data abstraction and event-driven visualization as important features of any visualization pipeline. One other technique first identifies links and directions of traffic flow and then transforms these attributes into HSI(Hue-Saturation-Intensity) color space to be mapped on the model [33]. Yin and Wolfson [32] proposed a weight-based map-matching method, which selects a most likely route for a vehicle. Their weighing is based on arcs and does not consider turns. Zhang et al. [34] studied a trajectory splitting model for efficient spatio-temporal indexing. Patel et al. [22] explored the efficient index for predicted trajectories. Visualization has also been implemented in several other public welfare services, ranging from a basic amenity of city pipeline [31] to a bigger picture of planning a city model [25]. Geographic visualization experts *itoWorld* [11] have developed time lapse animation of public transportation based on NPTDR data [3]. They use the time table schedule of each service from its origin to its destination and show the flow of buses, trains, and coaches over the course of day.

Visualization models: One of the interesting research efforts is to focus on visualizing the incident driven traffic data [26] to enable users to comprehend simultaneous events occurring during a particular traffic incident. Ning et al. [19] developed a visualization tool to aid traffic analysts to comprehend a traffic situation by a feedback chain model, where feedback is sent between traffic data, traffic flow detection, and traffic signal control. New visualization toolkit is developed to help end users to quickly build their customized visualization application without programming [24]. Guo and Wang [13] presented a trajectory visualization tool that focuses on visualizing traffic behavior at one road intersection. Ferreira et al. presented a system to help extract interesting patterns from the large trajectory data of the New York taxi trips [10], while Wang et al. [30] devised a visual analytic system to help understand the traffic congestion behavior in the City of Beijing using GPS trajectories. Recently, Chu et al. introduced a technique to transform the geographic coordinates (i.e. latitude and longitude) to street names reflecting contextual semantic information for the discovery and analysis of the hidden knowledge of massive taxi trajectory data within a city [8]. Different from the work on the analysis of large scale dynamic trajectory data, we aim to mine interesting patterns from the public transportation data that are essentially some network data. Particularly, we are interested in answering the quesiton of how the evolution of the transportation services over the years may indicate certain social changes in a large scale in space and time, such as the migration of the population and the change of the job opportunities.

III. Data and Tasks

In this section we introduce all the definitions which will be used in follow-up sections. We will first explain the Transportation Data of Great Britain, called NPTDR (National Public Transport Data Repository) [3], its format and structure. Next we explain the scope of our application in the form of tasks, where each task addresses a particular aspect of the transportation structure of a city.

A. NPTDR

NPTDR is a snapshot of route and timetable data for all public transport services in Great Britain. It is generated for a sample week every year starting in 2004. The week is normally the first or second complete week (Monday to Sunday) in October of each year, to ensure that it avoids school holidays or other seasonal variations and to achieve consistency from one year to the next.

NPTDR is complied from data supplied by 11 travel line regions of Great Britain, including Scotland, North East and Cumbria, North West, Yorkshire, Wales, West Midlands, East Midlands, East Angalia, South East, South West, and London. Each travel line region gets the data from its region's local transport authorities that compile information for bus, train, coach, and ferry services. Data from each travel line region are then collected into the central database of Great Britain as shown in Figure 1. Also, services which operate between points in two or more regions may be represented more than once in the national data. In our model we have used data for bus, train, metro, and coach services.

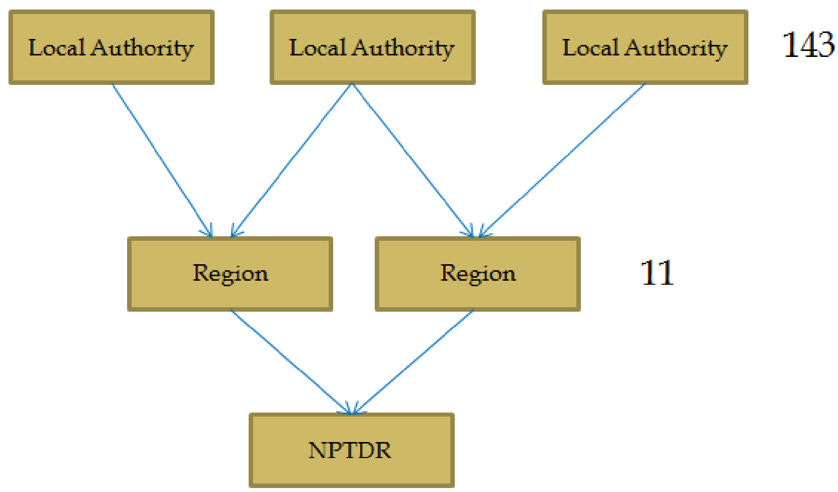

Fig. 1: NPTDR - 11 traveline regions, 143 local authorities

NPTDR routes and timetable data are referenced to stops coded in NaPTAN (National Public Transport Access Nodes), which in turn is referenced to NPTG (National Public Transport Gazetteer). The download of NPDTR [3] also includes the latest version of NaPTAN and NPTG.

The data supplied to NPTDR are of the format ATCO.CIF or TransXChange format. We use ATCO.CIF, which is a legacy format best described as position-specific data records. Figure 2 illustrates the structure of this format.

Note that the NPTDR data do not provide live service timing information; it has the fixed time schedule. This means that the services in real-time may not be the same as the time indicated in data set; i.e., a service may be early or delayed by few minutes based on the real-time conditions. We have taken this into account as a delay variable. Also, NPTDR does not have the information on service fares or number of people using the services. We are not considering these aspects in our application.

Below are the lists of definitions that are used throughout this paper.

1) **Service:** It is a mode of public transport, and it can be either bus, train, or coach.
2) **Journey:** It is the movement of a vehicle (bus, train, or coach) over a sequence of stopping points and the times from its origin terminal point to a destination terminal point (it is a column in the ATCO.CIF file).
3) **Stops:** It is the geographical locations at which the event happens during a course of **Journey**.
4) **Route:** It is comprised of a group of **Journeys** with more or less similar stopping points.
5) **Route number:** It is a unique identifier of each route.

B. Tasks

Given the above data, we are interested in helping the user answer the following questions.

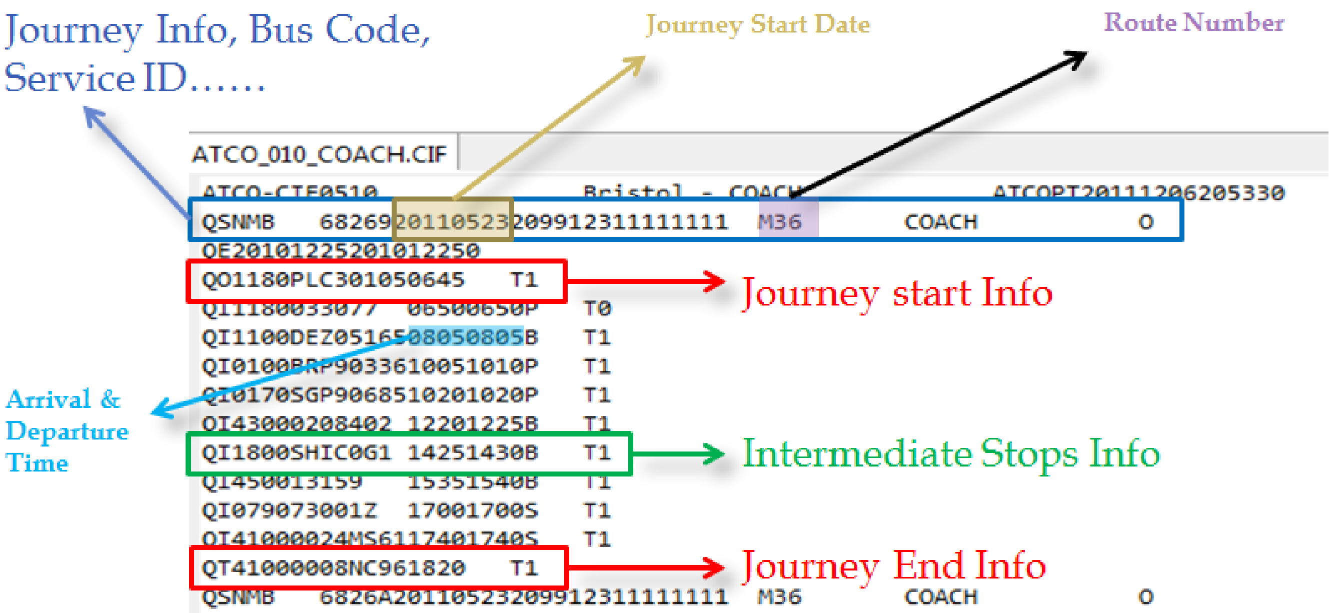

Fig. 2: ATCO.CIF data format

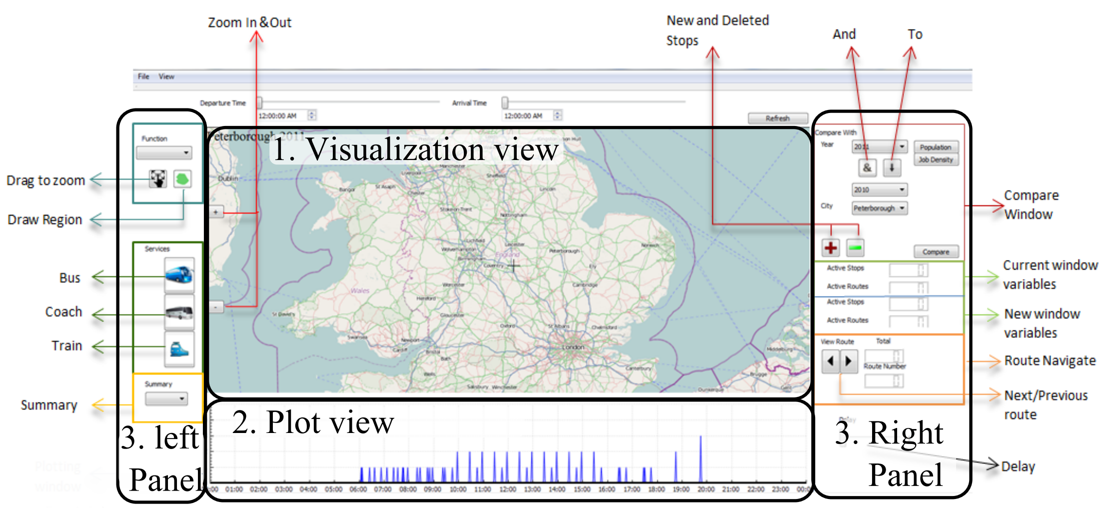

Fig. 3: User interface of our system.

Q1: *How many stops of a region are accessible at a given time range, say between 2pm to 3pm?* Note that some stops do not have services at certain time ranges, say between 11pm to 5am next day. Stops that have services in the late night may indicate the commercial attractions of the city.

Q2: *Which areas are the most (or the least) connected locations?* Note that this is different from **Q1** as it asks for the total services without a specific time range. Answering this will help us understand how far and wide the services from a city are available to other cities or locations.

Q3: *Given a time range, which regions (or stops) have a large number of services, e.g., many buses will stop by these areas? Do these regions with large number of services match the city areas with large job density?* Answering this question will help us identify the most busy places in a city. It may also be used to estimate the possibility of the occurrence of traffic congestion in a region at a given time instance due to certain unpredicted delay of the services. This information can help the traffic analyst evaluate the service quality at the individual stops or within a local region.

Q4: *How the transportation connectivity within a city can affect its population and job distribution over the years, and vice versa?* Here we try to identify certain correlation of the increase (or decrease) of the population and job densities with the growth (or reduce) of the transportation services at a given region over the years. Finding this correlation will help the public transportation authority better plan their services for the future growth.

IV. System Design and Implementations

Our system is developed specifically for the visual analytics of longitudinal transportation data. It enables the end users to interact with a rich set of comparison queries, and provides flexibility in selection of the data ranging from a particular location, to a group of locations, and to an entire region of a city. This allows the end users to analyze the variations in the data over the different time-slices of the years (e.g., 2004 to 2011). In this section, we describe the design and implementation of our system.

In order to mine useful information from the raw NPTDR data, we first need to convert them so that our system can handle them properly. Specifically, we classify the raw data into two types: 1) polygon data (route data), and 2) point data (location data). We also need to convert the coordinate information of the services from the Easting-Northing coordinate system to the Latitude-Longitude coordinate system. In order to visualize the population [16] and job density information [15], we need to extract the boundary data of all the regions inside a given city. We use Open StreetMap [4] for the geospatial visualization. Open StreetMap (OSM) has shape-file for each country, city, and region [7]. We use these data to draw regions of a city, and allow users to interact with them. We have used the OSM data from MapIt website [1], which has all the boundary polygon information from the shape-files of OSM. More details about this data-preprocessing can be found in the supplemental material. Next we will describe the interface and functionality of our system.

A. User Interface

Figure 3 shows the User Interface (UI) of our system. It consists of the following widgets.

1) **Visualization view:** This view shows the geo-graphical information including the map, region boundaries, service routes and stops, and other information. It also supports to multiple views for the comparative visualization of the traffic data from different years. This can be achieved by selecting from the options under the *view* menu.
2) **Plot view:** This view displays a number of plots for the mining of different information.
3) **Right control panel:** It consists of: **1) Compare window:** Tools in this window compare the transportation data between different years. The *And* and *To* buttons are selective between years, where the *And* button compares the two years and the *To* button compares years within a given range. The UI window displays the population and job density of a city, by clicking on their respective buttons in this window. The *New* and *Delete* buttons identify all new stops added or deleted between two years. **2) Window variables:** The two window variables represent two different years. They show the number of active stops and the number of active services, at any point during the user interaction process. **3) Route navigator:** The route navigator window, aids the users to explore the individual journeys of a route passing through a location or a group of locations. This interaction also triggers the plotting window and plots time schedule of a service passing through the stop(s).
4) **Left control panel:** It consists of 1) a *Function* drop-down list that provides support to the user queries applied to the selected area via the *Draw Region* button; 2) the *Services* sub-panel for the selection among bus, coach, and train; 3) a *Summary* drop-down list that displays the plots of the pre-computed information.

B. System Functions

Next, we demonstrate our visualization for each of the tasks as a functionality and the various ways the end users can interact with them. The implementation details can be found in the supplemental material. In the following demonstration, we use the transportation data for the city of Peterborough unless stated otherwise.

1) Functionality 1: Accessibility of Stops: Accessibility of Stops is defined as the availability of a stop at a particular time. More precisely, it implies that a particular service (bus, train, or coach) is at the stop at that time. We provide the functionality so that the end users can select a year between 2004 and 2011 in the *compare window*, and see the active stops by changing the time on the *arrival* and *departure* time sliders. This interaction also triggers current window variables, which show the number of active services and routes within the time frame.

In this example, the number of active stops are 879 and the number of routes passing through these stops are 3751. The blue dots indicate a stop. By moving the time slider, the user can choose different time intervals to inspect. In this example, we find that the locality *East* has the highest number of services running throughout the day, indicating it as an important region that might have good job facilities. [[Guoning: Can we highlight a region that has late buses?]]

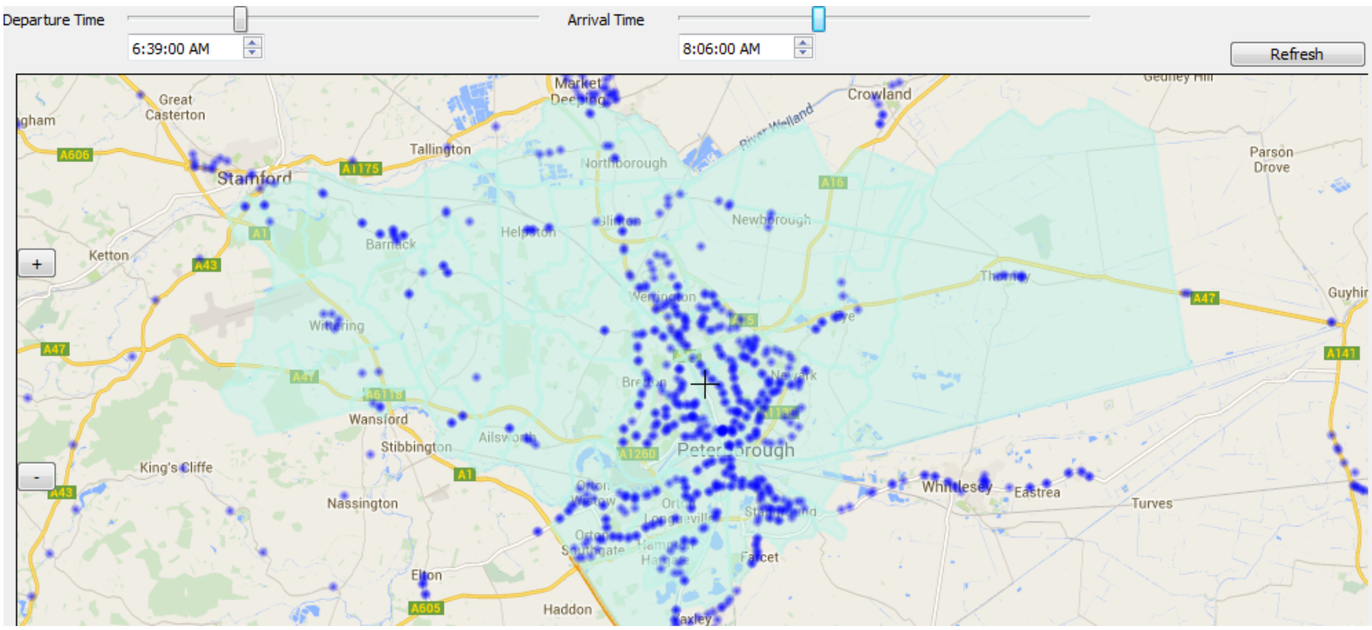

Fig. 4: Active bus stops between 6:31am and 7:39am.

2) Functionality 2: Connectivity Strength of Stops: Our system can compute and visualize the connectivity strength of all stops by calculating the number of routes passing through each stop and show its distribution. The users can select particular route strength, by clicking on the scalar color bar at the bottom of the screen as shown in Figure 5. This feature helps the users to identify patterns for particular route strength and then compare it between different years (Figure 9) or identify patterns between different route strengths of a year. From the visualization shown in Figure 5, we see that as the day progresses more routes become active, and towards midnight and the early morning route strength decreases. We also observe that the city of Peterborough has good service connectivity with the city Norwich and the strength in their connectivity is high. On an average, there are at least 20 stops with more than 50 routes passing through them from Peterborough to Norwich.

3) Functionality 3: Single Stop Analysis For Congestion Estimation: Our system enables the user to analyze the services provided at a selected stop. The user can select a stop from the visualization view, the profile of the services provided at the selected stop will be shown in the plot view. For instance, Figure 6(a) shows the plot of the bus services provided at a selected stop. From this plot, the route that provides the most services at that stop is easily revealed, so are the changes of the services of these routes over time.

Our system also supports the study of how a delay of certain bus affects other buses who schedules are nearby. In reality, given a time schedule of a route at a particular stop, the bus may arrive ahead or after the scheduled time. Depending on how many other buses will provide services at this stop, this discrepancy may or may not cause the traffic congestion at this stop. We study this by estimating the possibility of service overlap, i.e., how close it is between any two bus services of the same or different routes, in a given time range at the stop. This is because the larger the overlap is, the more likely the congestion may occur. This in turn can be utilized to decide whether the service arrangement in the given time range is appropriate or not. To achieve this, we assign a Gaussian function, $g_i(\tau) = exp(-\frac{(\tau-\mu_i)^2}{2\sigma_i^2})$, for each scheduled service at the specific stop, which describes the probability that a bus arrives the stop ahead or after the scheduled time μ_i by T minutes ($T < \sigma_i$). We then compute the 1D function $G(\tau) = \sum_{i=0}^{N} g_i(\tau)$ over the 24 hours period (Figure 6(b)), which we refer to as the *service overlap function*. This enables us to inspect the amount of the service overlap at the stop at any time. We also provide an interface, i.e., the *Delay* slider bar, to enable the user to control the delay time. With this functionality, we can estimate the quality

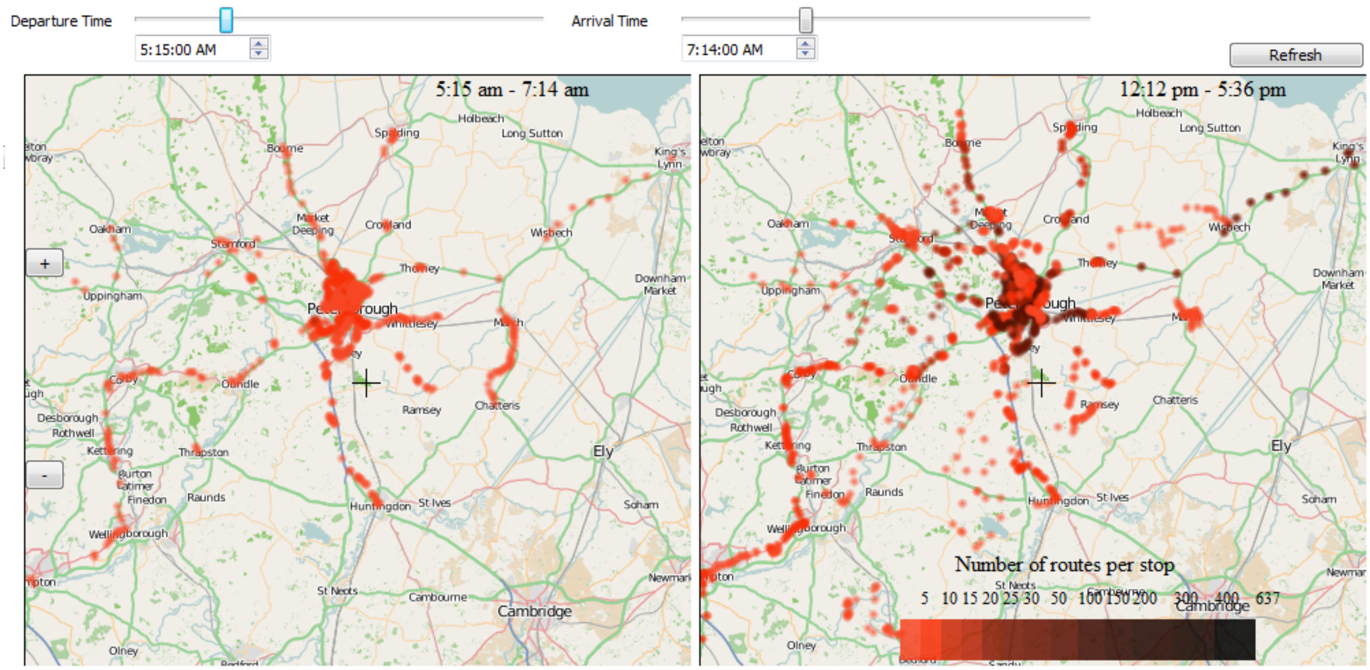

Fig. 5: Route strength at different times in Peterborough.

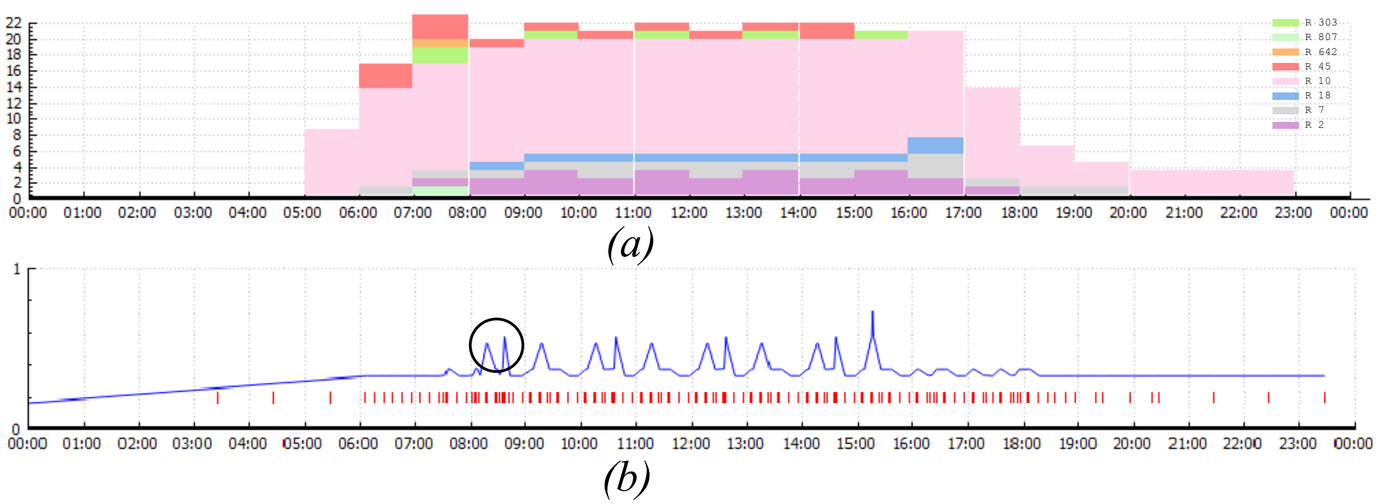

Fig. 6: (a) the bus services provided at a user selected top. Different color blocks show the numbers of buses of the respective routes that provide services at the stop. (b) the service overlap function at a selected stop.

of the services provided at the selected stop. We consider the quality is good if the overlap function is smooth with less spikes, while the services may not be ideal if its overlap function has many spikes (see the time period highlighted by the black circle in Figure 6) , because these spikes may indicate a higher chance of getting into congestion around the stop. Information like this can be used to improve the service schedule at the individual stops in the future. One can further extend this for the estimation of the potential congestion within different city regions at different given time. This can be achieved by estimating the density of the buses within any city regions using a 2D symmetric Gaussian function for each individual bus given its current service location. The higher the density, the more likely it will lead to traffic congestion.

4) *Functionality 4: Population and Job Density:* Our system visualizes the population and job density distribution on top the city map. Figure 7 shows the population density for all regions in Peterborough. The population density of some regions are not available. These regions are mainly towards the south of Peterborough. The population

density shown in the figure scales from white to red, where white indicates least populated regions and red indicates the highest populated regions. The regions towards the center have no population data available (displayed in white with no boundary); however, we suspect they have high population density since the services in this region are high. In Figure 7, we also visualize the active services over the population density to enable the user to inspect the correlation between the intensity of the services and the density of the population.

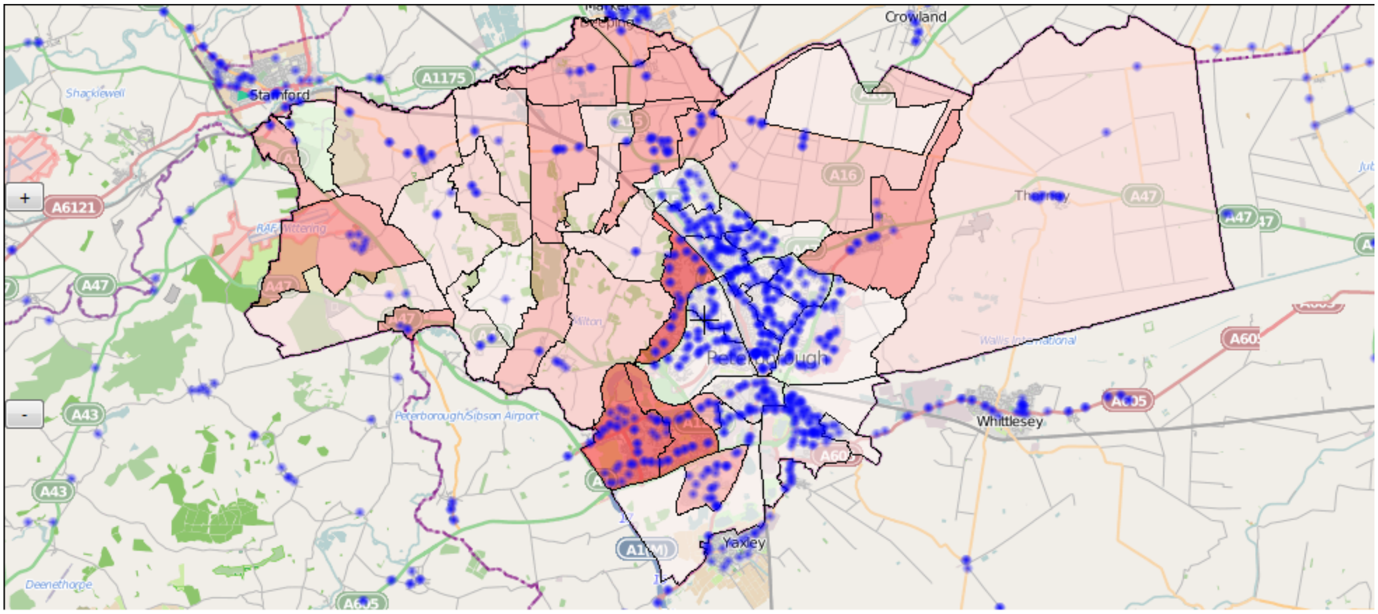

Fig. 7: Services shown over population density in Peterborough.

The blues circles shown in the above figures are interpolations around the center with half transparency. We use this to differentiate dense locations from sparse locations. For example, consider a location having 20 stops within a mile radius and another location having one stop within a mile radius. In both the cases, if the map view allows one-mile radius as a pixel size, both these locations will look the same, but by using interpolation and transparency we make one pixel darker and the other lighter.

From Figure 7, we observe that the region *Bretton* has the highest population, which is 60 times higher than the average population density in Peterborough. However, the availability of services in Bretton is low as compared to other regions; in fact, it is less than the average. This either means that the people in Bretton commute to the other regions for work or Bretton has poor transport services. OrthonLongueville and OrthonWaterville are the only two regions that have an ideal balance between population and service strength.

5) Functionality 5: Comparison Over the Years: Our system supports the comparison of the transportation data and the population/job density from different times or years in the *Visualization view*. Specifically, the user can open a new window using *view* menu item and view the active services for two different time periods. The *current window* and *new window* variables show the respective counts for active stops and services within the time frame. With this functionality, the aforementioned information can be compared over time to help us find interesting trends.

For instance, Figure 8 shows the comparison of the services between 5:31am to 6:56am for the year 2005 (left) and 2010 (right), respectively. In this example, we can see the difference between these years, the number of active services between 5:31am and 6:56am increased from 2005 to 2011. In 2005, there were 455 stops, 2010 has 562 stops, and 2011 has 571 stops. Also, there is a change in the connectivity of bus services in Peterborough. In 2005, the town Thorney had no services and the town Ramsey was connected between 5:31am and 6:56am. While in 2011 and 2010, Thorney has new services running in Peterborough and Ramsey is no longer connected.

Figure 9 shows the change of the connectivity strength between Peterborough and Norwich over the years. From this visualization, we see that in the past eight years the service connectivity and strength has changed a lot. From 2005 to 2008, Peterborough and Norwich had good connectivity of at least 200 or more routes running through each stop between them. However, there has been a sudden decrease in service connectivity in 2009. Since 2009, Peterborough and Norwich had no stops that provided 200 routes between them, the number of stops between them

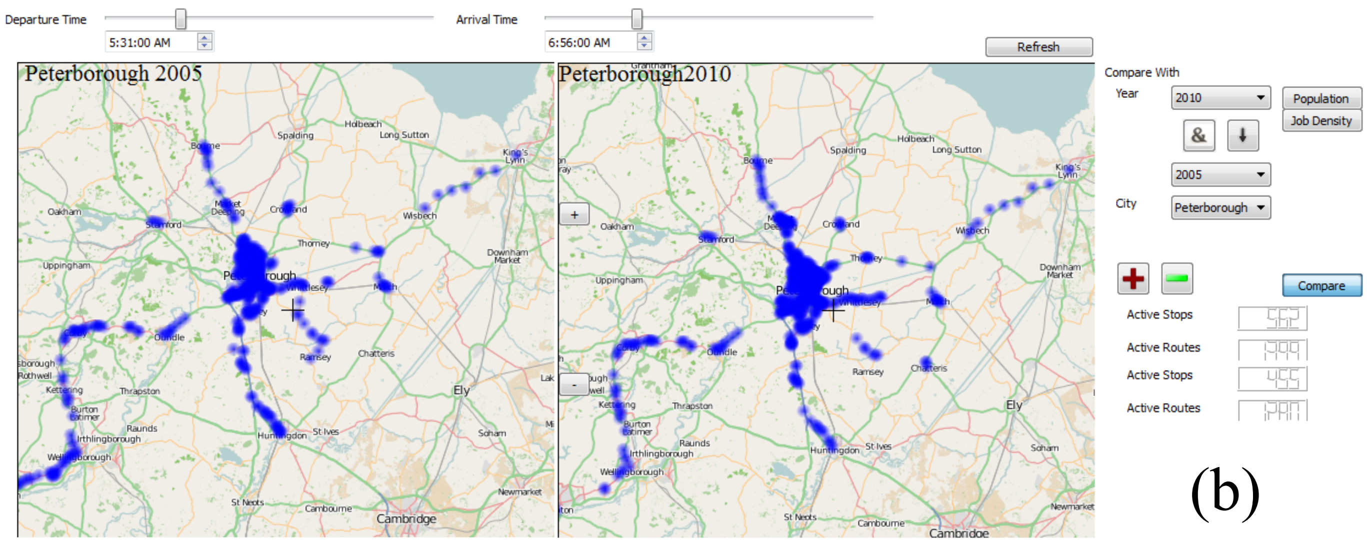

Fig. 8: Stops between 5:31am to 6:56am for the year 2005 and 2010.

decreased with each stop providing around 50 routes.

Figure 10 shows the service change of a city block of Peterborough between 2004 and 2011. The red dots indicate new stops in 2011 and green indicate deleted stops in 2004. The blue dots indicate the unchanged stops. Some points have both red and green indicating; these stops are renamed from prior years. This visualization shows that Peterborough had gone through a lot of change from 2004. Many stops have been added, removed or renamed.

V. Results and Case Studies

A. Case Study 1: City of Peterborough

In addition to the findings discussed in Section IV-B, we compare the transportation service strength with job density and population density of Peterborough in Figure 11. We observe that the population has increased consistently, despite the fluctuation of services. From years 2004 to 2008, the job density was dropping, which might cause the decrease of the transportation services. Since 2008, the situation of the employment seems getting better. However, the transportation services saw only a small increase then dropped again. This may indicate that before 2008, the residents of the city of Peterborough relied heavily on the public transportation to commute between jobs and homes. A detailed exploration enables us to find that from 2004 to mid-2007, Peterborough had more number of stops than routes passing through them. However, after mid-2007 until 2008, the number of routes has increased over the number of stops. This indicates that the number of people using the public transport has increased, thus making it busy. However, after 2008, they preferred using other ways to commute, which led to the reduce of services. This change may reflect the change of the life style of the residents of Peterborough. Another observation (not shown in this plot) indicates that after 2009, inter-city connectivity of Peterborough has increased, and as a result, there is an increase in the job density, which in turn attracts more people to settle down in Peterborough.

Figure 12 shows the service trend from 2004 to 2011. The train and coach services remain consistent, while the bus services fluctuate. Since job density has similar fluctuations as the bus services, we conclude that bus services characterize a closer look of a city's lifestyle, i.e., changes in bus services directly reflect the public preferences to commute, compared to train or coach services.

B. Case Study 2: City of London

In this case study, we explore the transportation system in London. We start by showing the general trend in service connectivity, population density, and job density from 2004 to 2011. We have also studied the metro service connectivity.

From Figure 13, we see that the public transportation services in London have increased since 2006. The increase from 2006 to 2007 was the steepest one. Since then, the service strength has remained approximately consistent,

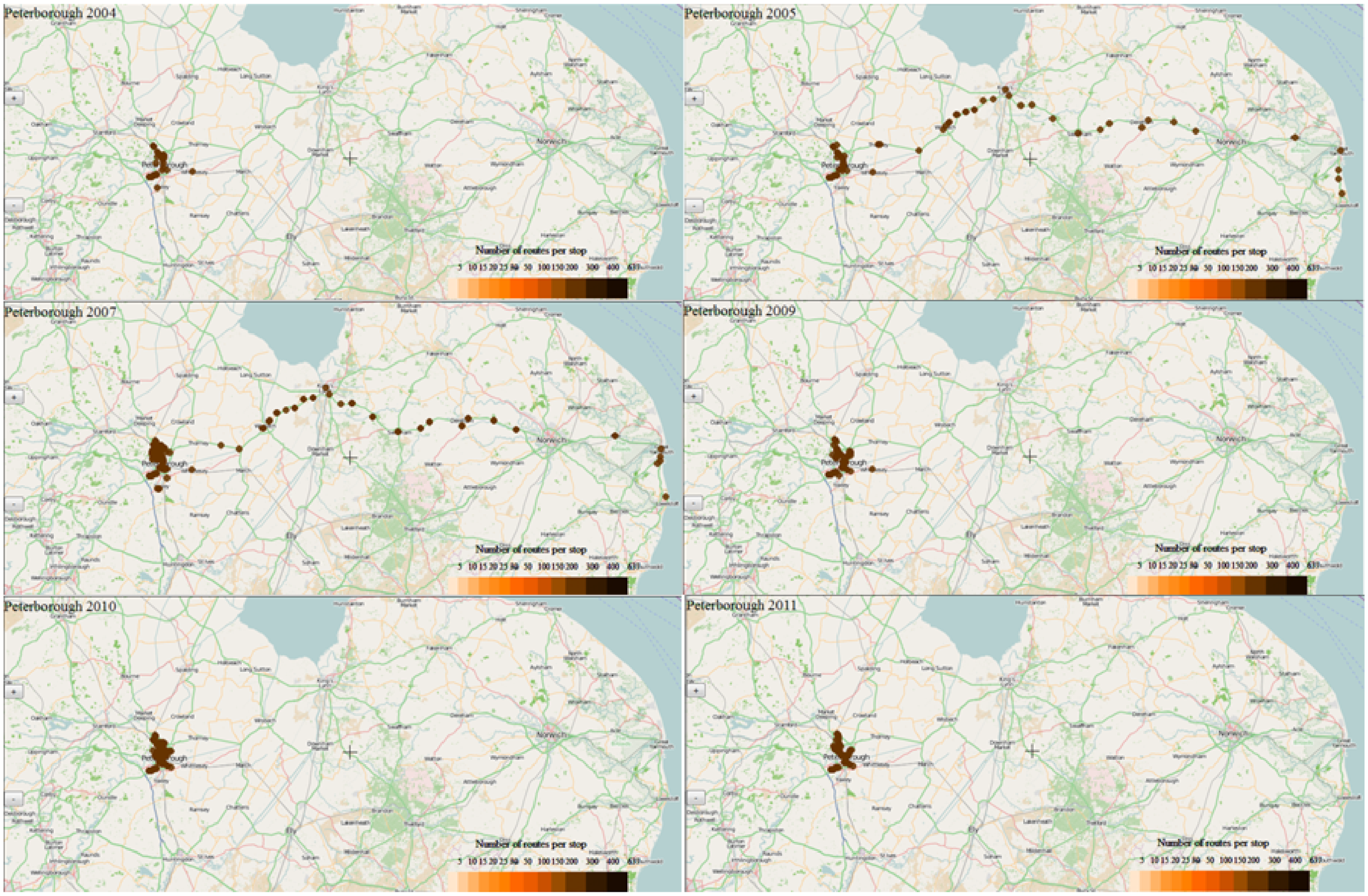

Fig. 9: Stops with more than 200 routes between Peterborough and Norwich over the years.

although it dropped a little after 2007. The job density increased from 2004 and reached its peak in 2006 and then decreased since 2006, reaching its minimum in 2009. Since then, the job density does not change much. The population density has increased consistently every year. From these observations, we deduce that service strength and job density change in a similar manner, and population density increased due to the more services in the areas. This is different from the observation of the city of Peterborough, where the decrease of the services did not prevent people from moving in. This is likely because London is a much bigger city with busy traffic. Most residents of London may prefer using public transportation than residents in Peterborough to commute. This indicates the different life styles of the residents between big cities and small cities.

The inset figure to the right shows the trend of the bus services and Metro services in London over the years. We see that the bus services have increased exponentially from 2006 to 2007, while the metro services remained consistent. This shows that the bus services contributed mainly towards the rise in service strength, as well as the increase of the population.

Using our application we have visualized this growth of the bus services in London from 2004 to 2011 (Figure 15). Interestingly, the visualization indicates that there might be other reasons behind this raise in service strength, when compared to what we observed from the graphs in Figure 13. One reason of this discrepancy may be the London 2012 Olympic Games. London won the bid in 2005, since then the bus services started growing.

Figure 14 shows the bus services in London between 2004 and 2011. It is clearly noticeable that there is a drastic

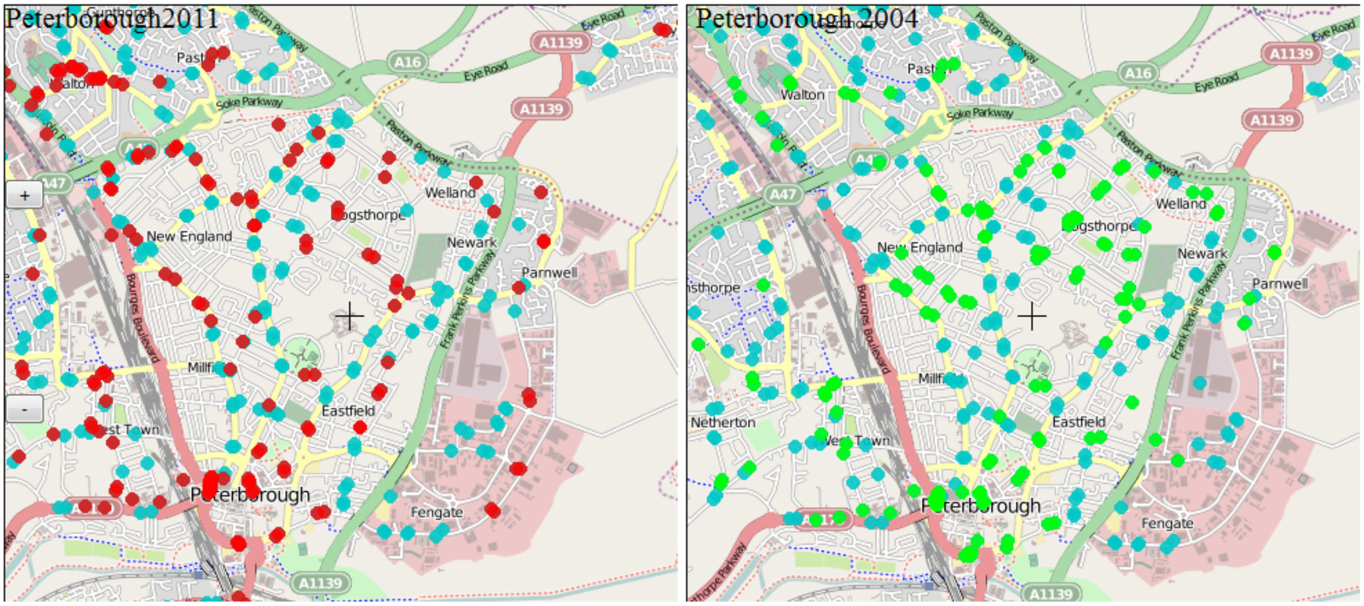

Fig. 10: New and deleted stops - 2011 and 2004 of a city block

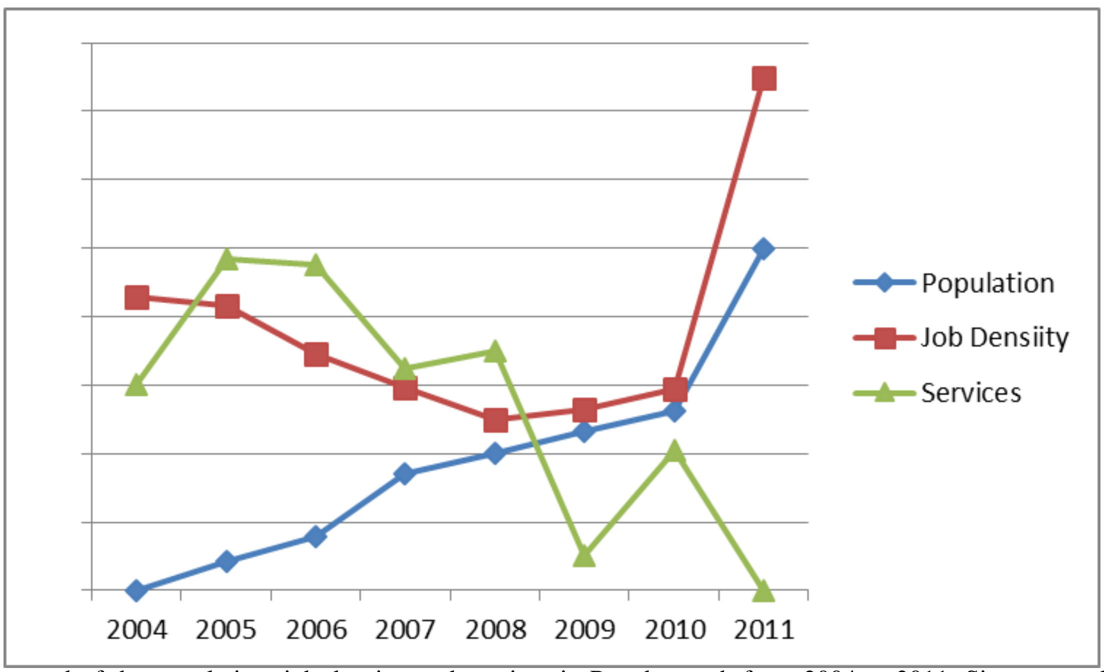

Fig. 11: The trend of the population, job density, and services in Peterborough from 2004 to 2011. Since we only care about the correlation of the trends of these three data, we normalize their values to facilitate the comparison.

change in bus services between the two years. Unusually, in 2004, most of the bus services were around at the boundaries of London and at the center of London (City of London), that means people were using other means to commute between these two areas. In 2011, we see that the bus services have increased and spread across London, this relates to the rise in population from 2004. It also indicates that more people have started to settle in London, and use bus services to commute to the City of London, we validate this pattern in the Figure 16. This is because, though the job density decreased in 2011, it mainly got concentrated at the center of the City of London; thus, allowing people to settle in other regions and commute to the City of London for jobs.

In Figure 15, we present the visual representation of the increase in the bus services of London. Every year since 2004, we see that the bus services increased within London, and in 2007, it increased exponentially compared to

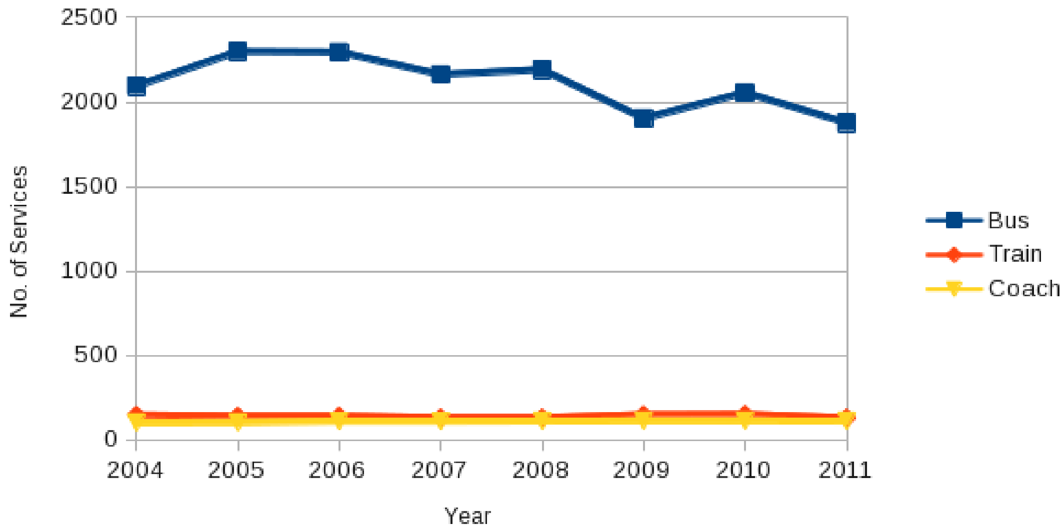

Fig. 12: Peterborough services from 2004 to 2011.

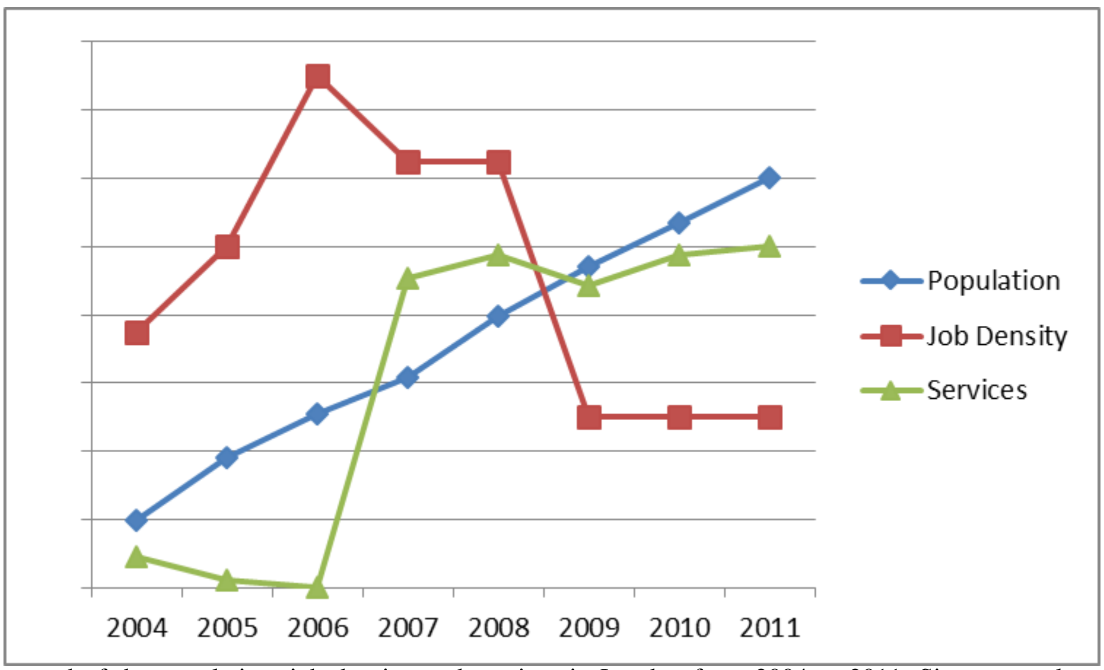

Fig. 13: The trend of the population, job density, and services in London from 2004 to 2011. Since we only care about the correlation of the trends of these three data, we normalize their values to facilitate the comparison.

2006. This happened because, although the overall job density reached its a new high in 2006, in reality the job density also increased drastically at the center of London. This made people to settle across London and travel to center of London for jobs. Figure 16 shows the job (top) and population (bottom) density of London in 2011. In this visualization, darker color indicates larger density values, while lighter color indicates smaller values.

C. *Case Study 3: Great Britain over the Years*

In this last case study, we extend our application to visualize the transportation strength of all the cities in Great Britain. We also visualize how every city differs from others in terms of the population and job distribution.

In Figure 17, the image on the left shows the population distribution and the image on the right shows the job distribution for the year 2011. Cities like London, Lincoln, Ripon, and Preston have good population and job

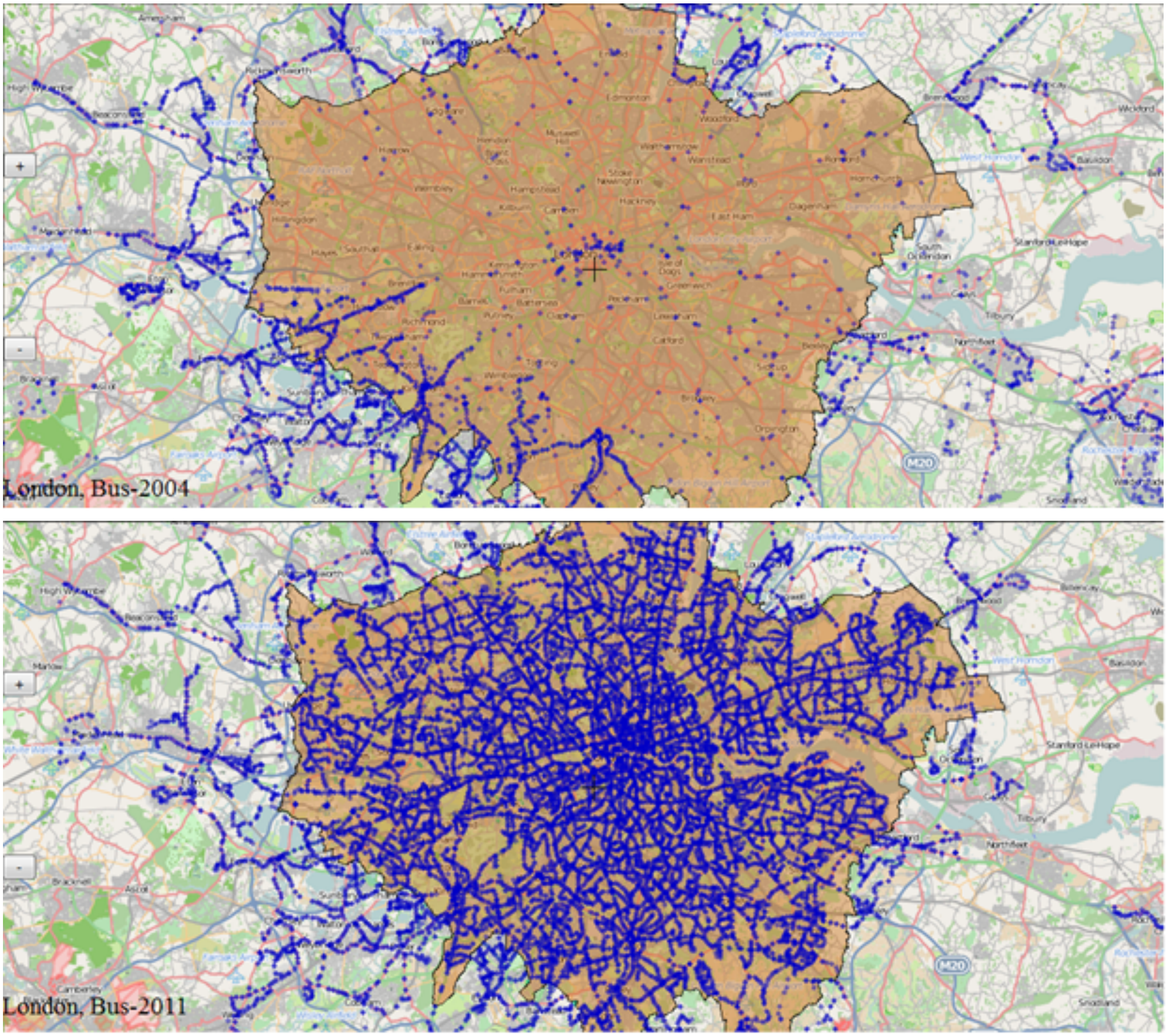

Fig. 14: Bus service of London in 2004 and 2011.

opportunities. While cities like Peterborough and other smaller cities around London have better job opportunities, though the population in these cities are less. This visualization gives a quick overview for all the cities, and is better than traditional bar chart representation. Since, we understand the geographical position of a city and at the same time its population or job density, giving us a wider scope for interpretation.

Figure 18 shows the coach and train connectivity from London to all the other cities for the year 2011. This visualization shows to what cities London is connected to, and at the same time, it shows how well is the service connectivity between them. For example, London and Bedford are connected through both train and coach services, and from the figure we can deduce that the coach services are more prominent over train services between the two cities (i.e., with darker red color). In the same way, when we compare London and Peterborough we see that they have only train services running between them. This visual representation helps us to understand other patterns, like the train services from London, connecting many cities compared to coach services. Interestingly, London does not have any train or coach services to its neighboring cities, namely, Chelmsford (Essex), Surrey, and Kent.

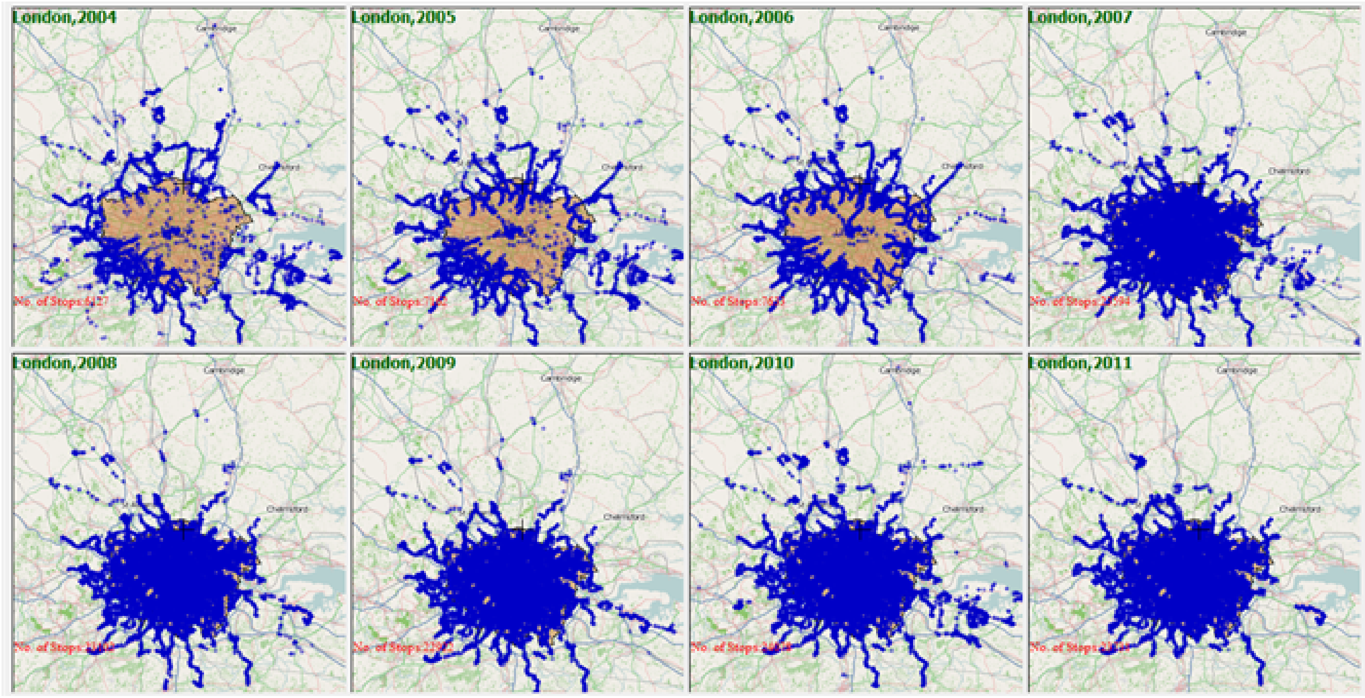

Fig. 15: Bus service of London from 2004 to 2011.

VI. Conclusion

We design a system for the visual analysis of the longitudinal transportation data of Great Britain. Through our system, we were able to successfully analyze different types of data together and show the inherent dependency between them, which might otherwise be difficult to observe. We have analyzed various features of the transportation data, over different aspects of city geography. From the study of both Peterborough and London, we saw some similar trends. In both the cases, the bus services seem to fluctuate more, compared to other services. Our analysis over other cities revealed a similar pattern. Therefore, we conclude that the bus transport expresses a city's dynamics and might directly affect the job and population distribution. Other transport services like train, coach, and metro show the general connectivity strength of a city, but do not point out the city subtleties.

Another interesting observation we made from the service trend, from 2004 to 2011 in both Peterborough and London is that in both the cases, first, there was a change in job density, then a similar change followed in the transportation service strength. That said, the development of the public transportation is always behind the demand of jobs. To utilize this observation, in order to increase jobs at a location, we might first need to focus on improving the transportation connectivity of that location.

References

[1] City boundary coordinates of uk at. http://mapit.mysociety.org/areas/UTA.html. IV

[2] Journey planner application. http://www.moovitapp.com/. II

[3] National public transport data repository data. http://data.gov.uk/dataset/nptdr. I, II, III, III-A

[4] Open street maps. http://wiki.openstreetmap.org/wiki/Main_Page. IV

[5] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *Visualization and Computer Graphics, IEEE Transactions on*, 14(1):47–60, 2008. II

[6] N. Andrienko, G. Andrienko, and P. Gatalsky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, 14(6):503–541, 2003. I

[7] G. B. boundary shapefiles. http://download.geofabrik.de/europe/great-britain.html. IV

[8] D. Chu, D. A. Sheets, Y. Zhao, Y. Wu, J. Yang, M. Zheng, and G. Chen. Visualizing hidden themes of taxi movement with semantic transformation. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 137–144. IEEE, 2014. I, II

[9] P. Compieta, S. D. Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi. Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing*, 18(3):255 – 279, 2007. I

[10] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2149–2158, 2013. I, II

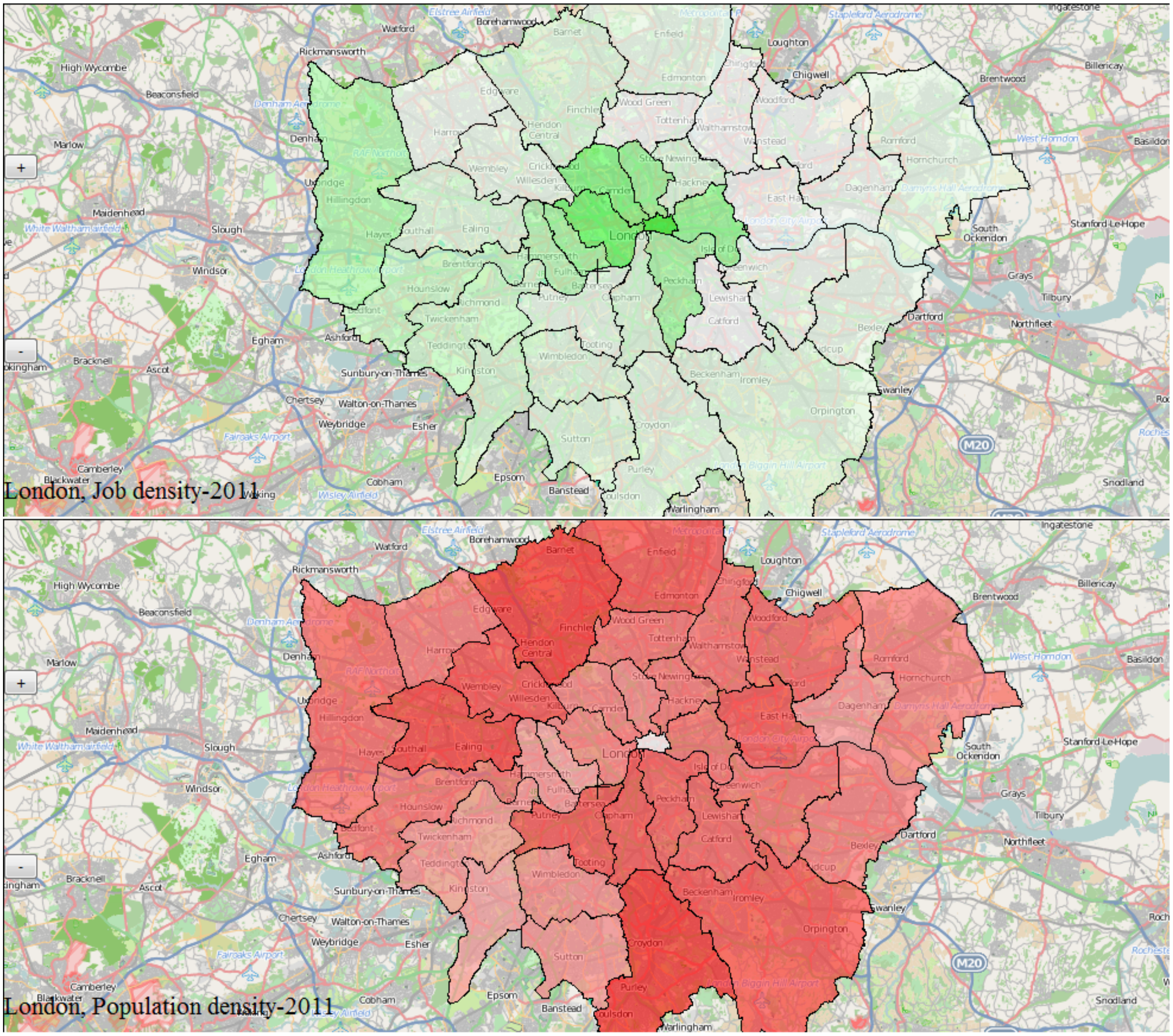

Fig. 16: Job (top) and population (bottom) density of London regions in 2011.

[11] T. flow visualization of great britain from vimeo. II

[12] O. Franzese and S. Joshi. Transportation applications of simulation: traffic simulation application to plan real-time distribution routes. In *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, WSC '02, pages 1214–1218. Winter Simulation Conference, 2002. II

[13] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, pages 163–170, 2011. II

[14] A. B. Habtie. Cellular-cloud integration framework in support of real-time monitoring and management of traffic on the road: the case of ethiopia. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, MEDES '12, pages 189–196, New York, NY, USA, 2012. ACM. II

[15] J. D. in UK data. http://www.nomisweb.co.uk/articles/649.aspxl. IV

[16] P. D. in UK data. http://www.nomisweb.co.uk/articles/676.aspxl. IV

[17] T. S. Kam, B. Ketan Dileep, J. H. Tan, et al. Divad: A dynamic and interactive visual analytical dashboard for exploring and analyzing transport data. *World Academy of Science, Engineering and Technology Journal*, 71:834–840, 2012. I

[18] S. Liu, J. Pu, Q. Luo, H. Qu, L. Ni, and R. Krishnan. Vait: A visual analytics system for metropolitan transportation. *Intelligent Transportation Systems, IEEE Transactions on*, PP(99):1–11, 2013. II

[19] X. Ning, Z. Li, and Y. Zhang. A practical research on visualized spatial analysis of traffic volume data. In *Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE*, volume 1, pages 172–175 vol.1, 2003. II

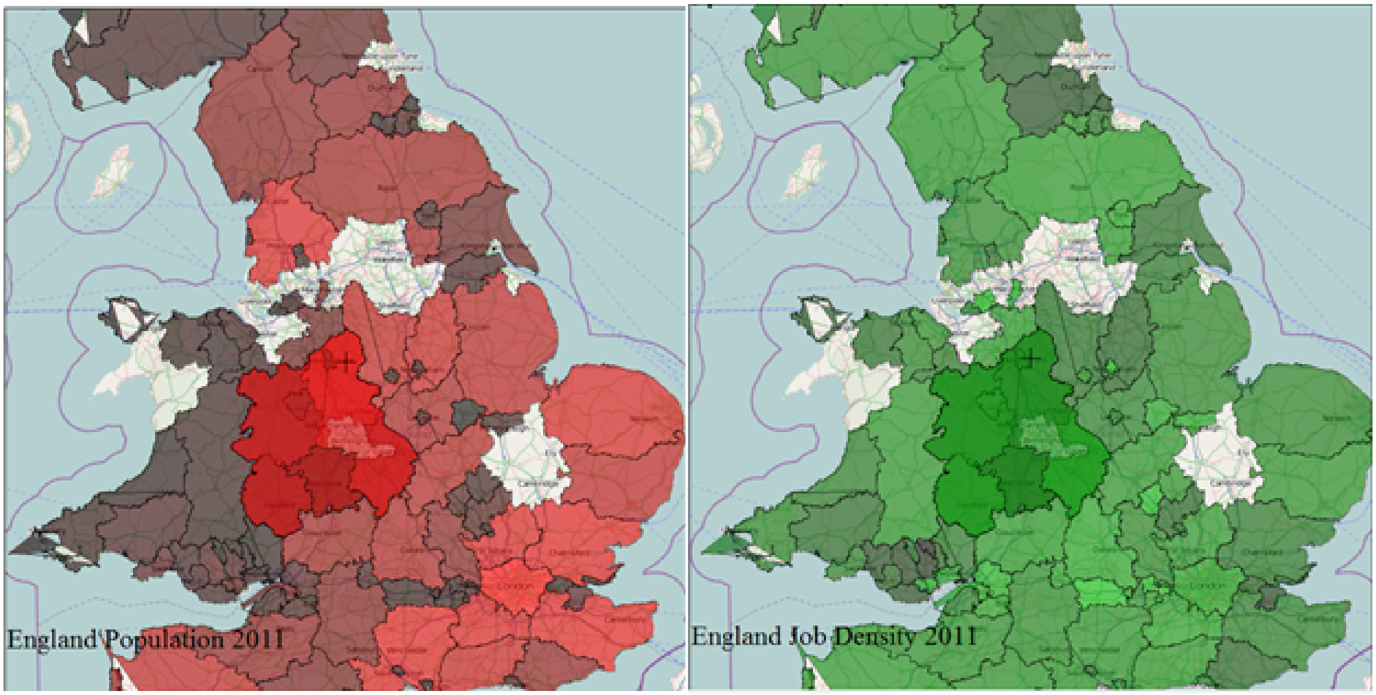

Fig. 17: England population (left) and job (right) density distribution in 2011. Some regions are left blank due to the lack of the data.

[20] G. B. online reservation. I

[21] M. L. Pack, K. Wongsuphasawat, M. VanDaniker, and D. Filippova. Ice–visual analytics for transportation incident datasets. In *Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on*, pages 200–205. IEEE, 2009. I

[22] J. M. Patel, Y. Chen, and V. P. Chakka. Stripes: an efficient index for predicted trajectories. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, SIGMOD '04, pages 635–646, New York, NY, USA, 2004. ACM. II

[23] E. Peytchev and C. Claramunt. Experiences in building decision support systems for traffic and transportation gis. In *Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems*, GIS '01, pages 154–159, New York, NY, USA, 2001. ACM. I

[24] L. Ren, F. Tian, X. L. Zhang, and L. Zhang. Daisyviz: A model-based user interface toolkit for interactive information visualization systems. *Journal of Visual Languages and Computing*, 21(4):209 – 229, 2010. II

[25] J. Royan, P. Gioia, R. Cavagna, and C. Bouville. Network-based visualization of 3d landscapes and city models. *Computer Graphics and Applications, IEEE*, 27(6):70–79, 2007. II

[26] M. VanDaniker. Visualizing real-time and archived traffic incident data. In *Information Reuse Integration, 2009. IRI '09. IEEE International Conference on*, pages 206–211, 2009. II

[27] M. Vasirani and S. Ossowski. A market-inspired approach to reservation-based urban road traffic management. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 617–624, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems. I

[28] K. M. Vaughn, M. A. Abdel-Aty, and R. Kitamura. A framework for developing a daily activity and multimodal travel planner. *International Transactions in Operational Research*, 6(1):107 – 121, 1999. II

[29] G. Wang, X. Shen, and H. Jiang. Research on growth trends and spatial distribution of shanghai population based on gis. In *Geoinformatics, 2010 18th International Conference on*, pages 1–4, 2010. I

[30] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. v. d. Wetering. Visual traffic jam analysis based on trajectory data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2159–2168, 2013. I, II

[31] D. Xiao, Y. Xiao, H. Lin, X. Fu, L. Xu, and J. Yang. A service stack for 3d visualization of gis based urban pipe network. In *Information Technology and Applications, 2009. IFITA '09. International Forum on*, volume 1, pages 342–346, 2009. II

[32] H. Yin and O. Wolfson. A weight-based map matching method in moving objects databases. In *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*, pages 437–438, 2004. II

[33] G. Zhang, J. Hu, R. Ma, Y. He, and Y. Zhang. Research on urban traffic and dynamic revolution based on visualized model. In *Computer Science and Information Engineering, 2009 WRI World Congress on*, volume 2, pages 70–75, 2009. II

[34] W. Zhang, J. Li, and W. Zhang. Spatio-temporal pattern query processing based on effective trajectory splitting models in moving object database. In *Computer and Computational Sciences, 2006. IMSCCS '06. First International Multi-Symposiums on*, volume 2, pages 540–547, 2006. II

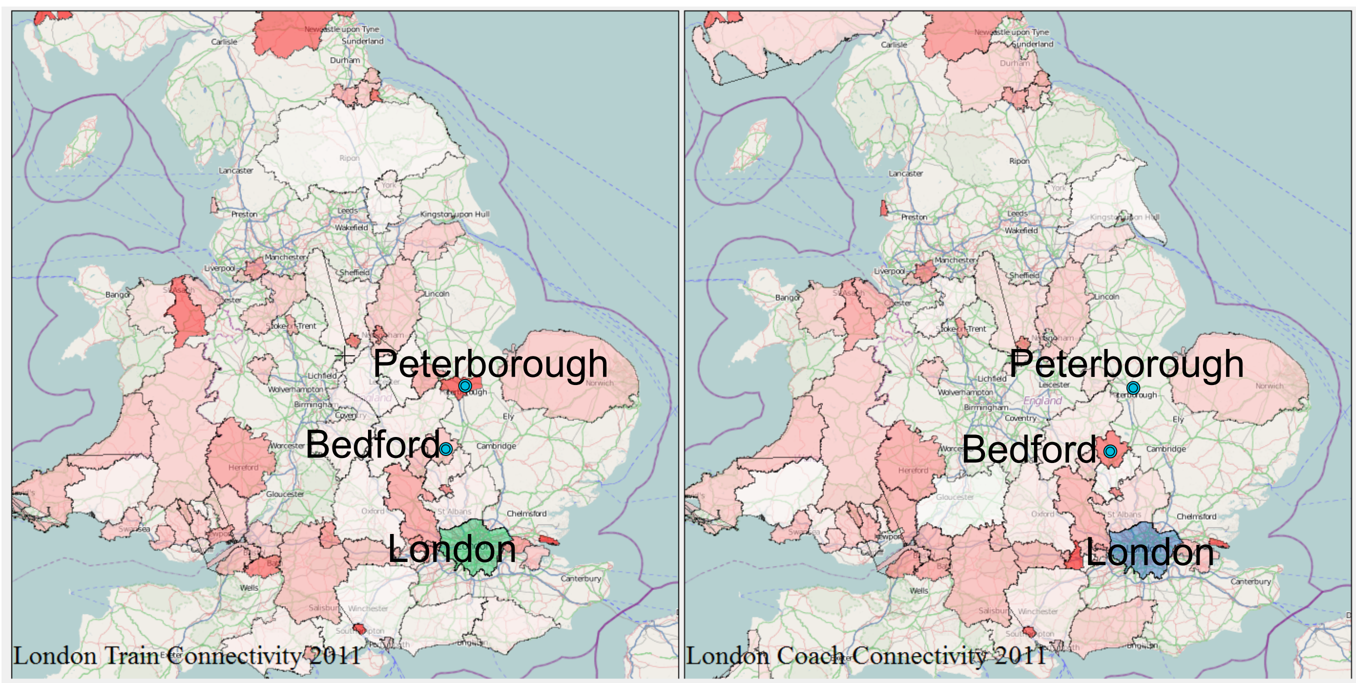

Fig. 18: London coach (left) and train (right) connectivity in 2011. The darker color indicates more services between London and the corresponding areas, while the lighter color indicates fewer services.