Bus Lines Explorer: Interactive Exploration of Public Transportation Data

Rainer Splechtna
VRVis Research Center
Vienna, Austria
Splechtna@VRVis.at

Alexandra Diehl
University of Buenos Aires
Buenos Aires, Argentina
adiehl@dc.uba.ar

Mai Elshehaly
University of Maryland
Maryland, MD, USA
mai@umbc.edu

Claudio Delrieux
CONICET and U.N.S.
Bahía Blanca, Argentina
cad@uns.edu.ar

Denis Gračanin
Virginia Tech
Blacksburg, VA, USA
gracanin@vt.edu

Krešimir Matković
VRVis Research Center
Vienna, Austria
Matkovic@VRVis.at

ABSTRACT

Public transportation movement data provide a wealth of information and insights into many aspects of urban life and human behavior. However, huge amounts of raw data, coupled with incomplete or inconsistent records, may turn into an obstacle for the effective use of the available information. The need for effective movement data analysis has resulted in a large number of visual analytics tools and specialized views. There are still many challenges in public transportation and other kinds of cyclic movement data analysis. In this paper we address some of those challenges by presenting an improvement of the standard map view. This improved view is specifically designed to simplify and make the visual analysis of complex movement data easier to perform, especially when integrated in a coordinated multiple views tool and articulated together with other techniques. We illustrate the effectiveness of the view on public transportation data from Bahía Blanca, Argentina.

CCS Concepts

•**Human-centered computing** → **Visualization techniques;** *Visualization systems and tools;*

Keywords

Visual Analysis, Movement and Public Transportation Data

1. INTRODUCTION

Recent technological advancements enable the collection and processing of large amounts of movement data. Due to several associated benefits, commercial vehicles, taxi cabs, and other surface transportation vehicles and devices are being manufactured or equipped with position tracking devices. This trend is becoming widespread, including also public transportation and city buses. The analysis of past public transportation movement data is helpful in many aspects, for example to predict waiting times, traffic jams, or overall transportation times, all of which may aid to develop potential improvement strategies [4, 10]. However, given the size, complexity, and low integrity of the associated datasets, analyzing the information is far from trivial. Besides automatic methods, visual analytics is one of the most successful means to deal with complex movement data [1, 5, 11, 15].

In this paper we describe an improved standard map view supporting interactive item filtering developed to better suit the data characteristics and typical analysis tasks. The view is integrated into a coordinated multiple views (CMV) tool to provide an advanced visual analysis system, bus line explorer, for interactive exploration and analysis of city buses movement data.

Our goal is to effectively deal with data clutter. This includes switching the context on and off to better see lines in focus. We illustrate the use and advantages of the map view using the public city buses data from Bahía Blanca, Argentina. The public transport system in Bahía Blanca uses RFID cards for passengers.

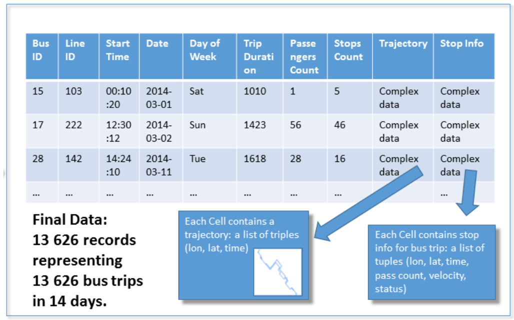
Every passenger has to register (check-in) using the card when entering a bus. It is not necessary to check-out when leaving the bus. The buses' GPS positions, along with some additional information, are registered and logged every six seconds. Finally, the station positions and vehicle status at stations (velocity, doors opened/closed, time spent in a station, etc.), are also recorded. Our visual analytics system takes full advantage of these multiple data sources in a comprehensive and intuitive way.

The main contributions of this paper can be summarized as follows: *1)* An improved standard map view with advanced filtering and brushing to deal with uncertainties in recorded locations; and *2)* Integration of the map view within a CMV tool to support advanced brushing and filtering.

2. RELATED WORK

Andrienko et al. [1] provide an overview of the main concepts and issues related to the analysis of movement data. They discuss the conceptual framework and transformation of movement data and then describe the visual analytics infrastructure and its use. Similarly, Pelekis and Theodoridis [12] discuss the characteristics of mobility data, mobility

VINCI '16, September 24-26, 2016, Dallas, TX, USA
© 2016 ACM. ISBN 978-1-4503-4149-3/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2968220.2968238



Bus ID	Line ID	Start Time	Date	Day of Week	Trip Duration	Passengers Count	Stops Count	Trajectory	Stop Info
15	103	00:10:20	2014-03-01	Sat	1010	1	5	Complex data	Complex data
17	222	12:30:12	2014-03-02	Sun	1423	56	46	Complex data	Complex data
28	142	14:24:10	2014-03-11	Tue	1618	28	16	Complex data	Complex data
...	...	...	...	...	...	...	...	...	...

Figure 1: The raw data from over 95,000 files was processed to create the final data table. Each trip is stored in a single record containing scalar attributes and two complex attributes (a list of tuples).

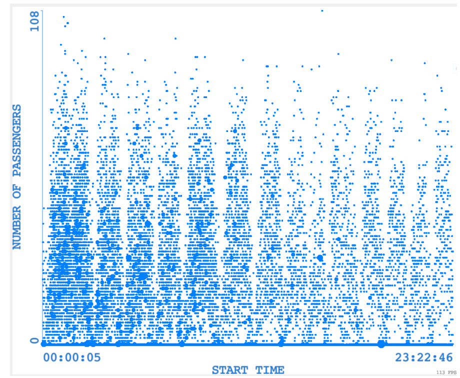


Figure 2: A scatter plot view showing the number of checked-in passengers during a bus trip ($y-axis$) based on the trip start time ($x-axis$).

data management, exploration, and semantic aspects. They also discuss using Moving Object Database (MOD) engines, including SECONDO and Hermes, to leverage spatial data warehousing and spatio-temporal databases.

Visual analysis of spatio-temporal urban data provides many opportunities for data-driven analysis. For example, taxi trips data [5] can provide insights into many aspects of city life and human behavior. Similarly, public transportation systems affect many aspects of urban life. These systems are very complex and, consequently, the collected data are very difficult to effectively visualize and analyze [15].

The need for effective traffic data analysis resulted in many visual analytics tools [3, 9] using various data filtering, processing and displaying methods for spatial, temporal, numerical and categorical attributes of traffic data.

Public transportation schedules are designed to provide optimal transportation services. However, real-world services usually deviate from the original planning. Collecting and analysis of historical data on line-, trip- and station-level can help detect usage patterns and problems [11]. Such data is often in a form of time series that span over hours, days, months, etc. Specialized views, such as calendar-based views [14], can help identify patterns. Those views are used as a part of CMV tools [13]. Composite brushing and visual effects are used to highlight data points across views [8]. Similarly, visual query systems are used to support querying spatio-temporal data [2].

3. DATA AND PROBLEM OVERVIEW

Actual timing of transportation services could differ significantly from the official schedule. Those discrepancies could be large, and may result in a significant waste of time and money. The reasons for discrepancies may be difficult to determine. Hence, collected real-time and historical data could be used to determine schedule problems or to provide real-time arrival information [6]. Bus schedules present an especially challenging problem due to the many factors that can affect the original planification. Instrumentation and sensors could provide necessary information [7].

3.1 Data Sources

We use the public city buses data for Bahía Blanca, Argentina to illustrate our approach. The passengers use RFID cards to register (check-in) when boarding buses equipped with GPS tracking devices. Several other events are manually logged. The events logging is done by a bus driver resulting in great variations of bus-stop positions in the data. Sometimes drivers trigger the stop event before or after the actual stop. The passengers should check-in using their card as they enter the bus, but in many cases the bus already leaves the stop while the passengers are still registering.

A simple look at the map with bus-stop positions from many trips of the buses traveling the same bus line immediately shows the problem. A solution could be to merge the related bus-stop positions and to map them onto the actual position. This would, however, hide the occurrence of inconsistent station logs, which is a potentially useful information source. Each station log (bus driver triggered) contains position (longitude and latitude) and additional bus information, for instance the bus velocity. The velocity is often nonzero (not expected at the bus-stop) and goes as high as 7 km/h. The relatively low velocity indicates that bus drivers do trigger logging regularly, but they could also improve their precision. This example clearly shows that public transportation data is doomed to be plagued with this kind of noise that is not known in advance, thus reducing the quality of the data and the effectiveness of naïve use of standard data analysis tools.

The logs of each bus trip are saved into several files. There is a trajectory file with the regular samples of coordinates and corresponding times. There is a file with information about checked-in passengers. For each check-in, there is time and position information, as well as number of passengers that checked in simultaneously. Finally, there is a file describing bus stops, with time and position data, as well as additional data such as bus velocity, and door status (open or closed), for example. The files are stored in folders corresponding to the hour when the trip starts. The size of raw data for 14 days is approximately 3 GByte (stored in more than 95,000 files).

3.2 Problem Statement

As an example of the effectiveness of the improved map view, we explore bus trips and analyze how buses are used. For example, are there large variations in load across lines, days of the week, or hours? We explore how such data can

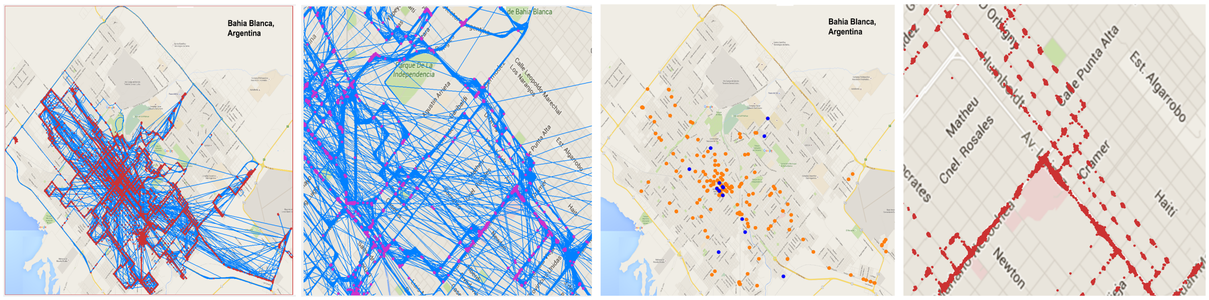


Figure 3: The map view can be configured in many different ways. The map on the left shows all trajectories (blue lines) and all stops (red points). If we zoom-in (the second map from left) we can see more details and unusual trajectories that do not follow roads. This is certainly worth further inspection. The third map from left shows only stops where more than 20 passengers checked-in. The map on the right shows a zoomed in view of bus stops only (red points). The trajectories are omitted to prevent clutter.

be used to plan bus schedules in order to deploy buses more efficiently while increasing the quality of service.

We create a data table where each bus trip has the corresponding record. A record contains scalar attributes and complex (structured) attributes — the trajectory attribute and the bus-stop attribute. The trajectory attribute contains a trajectory, which is a list of longitude-latitude-time triples. The bus-stop attribute contains a list of bus-stop events. Each bus-stop event is a tuple of bus-stop position, time, the time spent in stop, number of checked-in passengers, etc. Figure 1 shows the data structure (each record is represented as a row in a table). There are 13,626 records, each corresponding to a single bus trip.

This data structure supports easy visual selection of individual bus trips (records). A user brushes (interactively selects) records in a graphical view to select a subset of bus trips. For all standard, scalar attributes, standard visualizations can be used. If we want to show the total number of passengers on a bus trip (a scalar attribute) with respect to the bus trip starting time (another scalar attribute), we can use a standard scatter plot view and see if there is correlation (Figure 2). The size of points corresponds to the number of bus trips having the same value. Contrary to our expectations, there are many night trips with large point sizes. If we use a CMV tool we can efficiently drill down the data space in order to find some useful information. Since there are attributes that represent spatio-temporal movement data with events along trajectories, standard views are not sufficient. For this reason we introduce in Section 4 an advanced visual analysis system for interactive exploration and analysis of city buses movement data.

4. NOVEL VISUALIZATION

We improved the standard map view to support exploration and analysis of complex data. The improved map view supports some specific spatial tasks and custom filtering specific to our analysis. The view is fully integrated in a CMV tool. Many additional view types (scatter plot, histogram, parallel coordinates, etc.) can be used along the new view. However, the new view is always in the center of the analysis, as it directly shows the spatio-temporal data, makes it easy to establish a mental map of the data, and supports efficient analysis.

Visualizing aggregated values over time is not always sufficient for movement data analysis so we include the improved standard map view. As stated above, the data we are dealing with is collected during bus rides, and bus drivers trigger data logging manually. The bus-stop logs contain position, time spent in stop, velocity in stop, and bus status.

The analysts are also interested in log data. We improved the standard map view with advanced filtering. The improved standard map view can show bus lines and bus stops. The stops can be filtered in many ways. In addition to normal brushing in all views, we allow a user to filter bus stops based on the three aggregates used throughout the analysis: time at a stop, velocity at a stop, and number of passengers checked-in at a stop. As the analysis tasks are different, for each of the aggregates a threshold can be set. The user can select stops that have an aggregate value which is less than, greater than, or equal to a threshold.

We suggest several approaches to dealing with cluttered views. When a subset of data is brushed, the context can be turned on and off to help better see lines in focus. The underlying map can also be turned off. Finally, the map can be zoomed and panned. The zoomed in map reveals many details otherwise not seen. Figure 3 shows several of the above described features. The view on the left shows the map where all bus trajectories and all bus stops are shown. It is clear that the view is too cluttered. We can immediately spot unusual trajectories which do not follow roads. Zoomed-in view (the second view from left in Figure 3) confirms the first impression. The blue lines are drawn using GPS positions from buses trajectories. If a bus looses the GPS signal, or if the driver turns the device off, the last valid point will be connected with the next logged point, and the road will not be followed. The third map from left shows only the stations where more than 20 passengers checked-in. The trajectories are not shown to reduce clutter. The selection is refined using two brushes (orange and blue) and the calendar. The blue points show bus stops where more than 20 checked-in passengers during the weekend (Saturday and Sunday), and the orange points show the stops for the rest of the week. The map on the right shows a zoomed in view of bus stops only (red points). The trajectories are omitted to prevent clutter. All red points cannot be actual bus stops. The dispersion is due to the manual log triggering which happens at different positions.

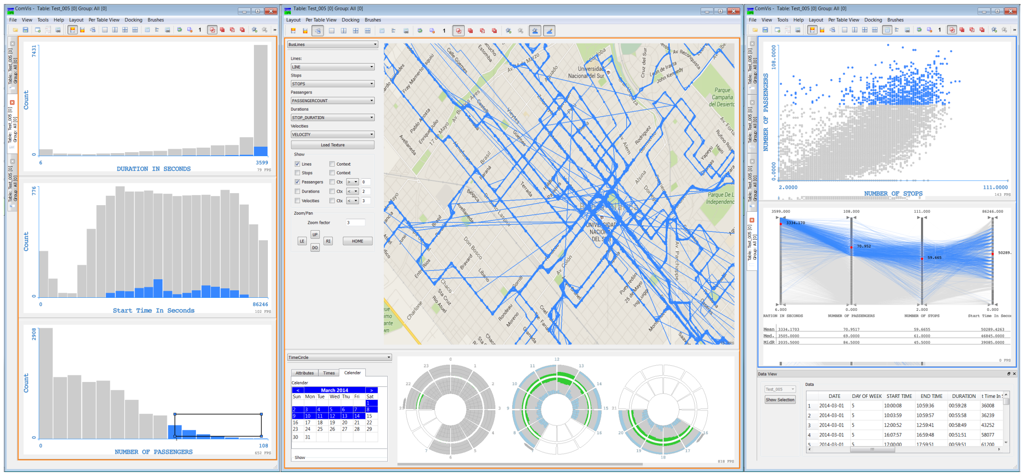


Figure 4: A screen shot of the CMV tool during an analysis session. The histogram views on the left depict several scalar attributes. The new view is shown in the center.

5. EVALUATION

In this section we briefly evaluate the improved map view using the two weeks data set for Bahía Blanca. We illustrate the usage of the bus line explorer system in two ways: data set quality exploration and bus lines exploration. These explorations illustrate typical tasks that a public transportation company or a policy maker might be interested in.

Many bus lines trajectories do not follow roads. This is especially true close to some bus stops. We have selected only bus stops where buses spent more than 200 seconds. The long time spent at a stop indicates that this is not a regular bus-stop. Actually this stops are garages, where buses go at the end of their schedule. It seems that bus drivers turn off GPS devices at the last station and turn it on when they arrive at the garage.

The station logs are another problem. The logging of the stations is triggered by drivers who might do it at different distances from the actual bus-stop. Consequently, all these points cannot be actual bus stops since some of them are less than a block apart. In some streets we can see clear clusters, and on the main avenue the stops are practically everywhere. When we got the data from the public transportation company, they told us that all bus stops are logged. We could not anticipate such dispersion of bus-stop locations. Automatic processing of such data would be very complex, while interactive visual analysis makes it possible to see the problems, and make the analysts aware of it.

Once the analysts better understand the data and potential problems, they can better asses complex dependencies hidden in the data. Figure 4 shows a screen shot during a typical analysis session. The new improved map view is in the middle. The three histograms on the left depict three scalar attributes: trip duration in seconds (between 6 and 3,599), start time in seconds (number of trips each hour during day), and a number of passengers who checked-in during a single trip (between 0 and 108). The scatter plot view on the right shows the number of passengers and number of stops. The parallel coordinates view on the right depicts four additional attributes. The table in the lower right corner shows all details for the selected trips. The user brushes trips with high number of passengers in the corresponding histogram and all views are updated. We can see that high number of passengers appears only from 21:00 to 22:00 during night, while it is quite frequent during daytime. This could be a reason to change the schedule. During the day, time interval from 12:00 to 13:00 is the busiest one. The map is often zoomed in and panned. The user can drill down further with additional selections, or by deploying filtering in the map view, or by constraining time using the calendar.

6. CONCLUSIONS

The analysis and assessment of movement data still represents a great challenge, especially when dealing with public transportation data. We addressed this challenge by developing a new view, the improved standard map view. The improved view supports easy, interactive filtering specifications to help data quality assessment. The view can be coordinated or articulated with other views within a CMV tool, resulting in a powerful visual analysis bus line explorer system. The effectiveness of the bus line explorer system was illustrated using public transportation data from Bahía Blanca, Argentina. Some data quality issues and inconsistencies were caused by the way how bus stops are logged, which would be a large obstacle for an automatic analysis. The bus line explorer system allowed us to successfully deal with those inconsistencies.

We plan to continue this research and to carefully evaluate the improved map view and the bus line explorer system with domain experts from public transportation companies. We also plan to provide semi-automatic methods for data processing/filtering during the first step in the analysis.

7. ACKNOWLEDGMENTS

Part of this work was done in the scope of the K1 program at the VRVis Research Center (http://www.VRVis.at). This work was partially supported by grants WTZ Argentina 2015-16, Project number AR 07/2015. We are grateful to the Secretaría de Gobierno Abierto of the Bahía Blanca city for providing the dataset of the public transportation.

8. REFERENCES

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, Berlin, 2013.

[2] V. M. B. Cavalcanti, U. Schiel, and C. de Souza Baptista. Querying spatio-temporal databases using a visual environment. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 412–419, New York, 2006. ACM.

[3] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984, Dec. 2015.

[4] D. Fairbairn. Geovisualization issues in public transport applications. In J. Dykes, A. M. MacEachren, and M.-J. Kraak, editors, *Exploring Geovisualization*, International Cartographic Association, chapter 26, pages 513–528. Elsevier, Oxford, 2005.

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

[6] B. Ferris, K. Watkins, and A. Borning. OneBusAway: Results from providing real-time arrival information for public transit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1807–1816, New York, 2010. ACM.

[7] S. Foell, G. Kortuem, R. Rawassizadeh, M. Handte, U. Iqbal, and P. Marrón. Micro-navigation for urban bus passengers: Using the internet of things to improve the public transport experience. In *Proceedings of the First International Conference on IoT in Urban Space*, pages 1–6, Brussels, 2014. ICST.

[8] A. L. Griffin and A. C. Robinson. Comparing color and leader line highlighting strategies in coordinated view geovisualizations. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):339–349, Mar. 2015.

[9] T. Nagel, M. Maitan, E. Duval, A. V. Moere, J. Klerkx, K. Kloeckl, and C. Ratti. Touching transport — a case study on visualizing metropolitan public transit on interactive tabletops. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pages 281–288, New York, 2014. ACM.

[10] M. L. Pack. Visualization in transportation: Challenges and opportunities for everyone. *IEEE Computer Graphics and Applications*, 30(4):90–96, July 2010.

[11] C. Palomo, Z. Guo, C. T. Silva, and J. Freire. Visually exploring transportation schedules. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):170–179, Jan. 2016.

[12] N. Pelekis and Y. Theodoridis, editors. *Mobility Data Management and Exploration*. Springer, New York, 2014.

[13] J. C. Roberts. State of the art: Coordinated multiple views in exploratory visualization. In *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV '07)*, pages 61–71, July 2007.

[14] J. J. V. Wijk and E. R. V. Selow. Cluster and calendar based visualization of time series data. In *Proceedings of the 1999 IEEE Symposium on Information Visualization(Info Vis '99)*, pages 4–9, 140, 1999.

[15] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu. Visualizing mobility of public transportation system. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1833–1842, Dec. 2014.