Visually Exploring Transportation Schedules

Cesar Palomo, Zhan Guo, Cláudio T. Silva *Fellow, IEEE* and Juliana Freire *Member, IEEE*

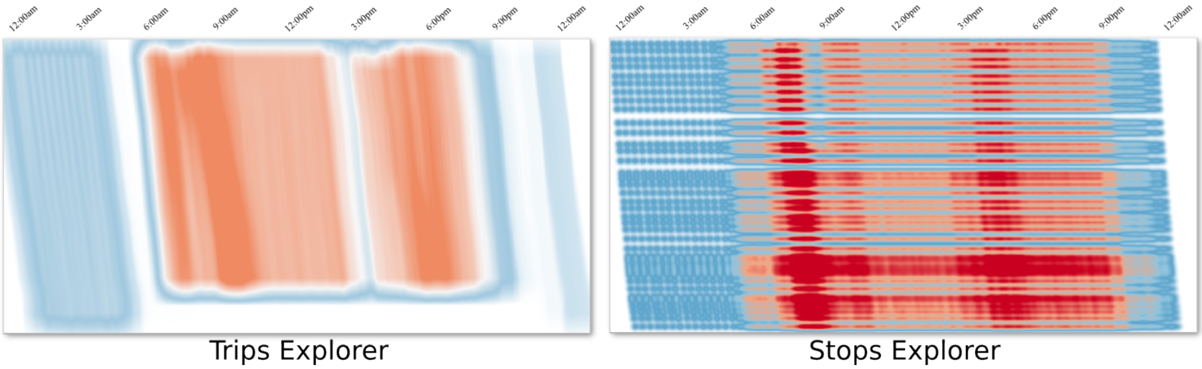


Fig. 1: TR-EX is a visual analytics system for detection, inspection and comparison of spatio-temporal patterns in transportation services. It uses and combines new visual encodings inspired by Marey's Graph to support the exploration of planned and real transportation schedules, showing where and when systemic or eventual deficiencies take place at trip- and station-level. The Trips Explorer highlights regions in time and space of low (in blue) and high (in red) frequency of trips. Here, periods of low frequency are observed at night and high frequency at peak hours. This visualization shows the different starting stations (in the vertical axis) and abrupt supply reduction (shown in white) before 3pm and before the night peak hours. The Stops Explorer allows exploration of the data associated with the stations in a route, including wait time and reliability. The figure shows how wait times vary throughout the day for stations along NYC subway line 1 in the northbound direction – wait times observed in Uptown stations are much longer than in the Downtown ones.

Abstract— Public transportation schedules are designed by agencies to optimize service quality under multiple constraints. However, real service usually deviates from the plan. Therefore, transportation analysts need to identify, compare and explain both eventual and systemic performance issues that must be addressed so that better timetables can be created. The purely statistical tools commonly used by analysts pose many difficulties due to the large number of attributes at trip- and station-level for planned and real service. Also challenging is the need for models at multiple scales to search for patterns at different times and stations, since analysts do not know exactly where or when relevant patterns might emerge and need to compute statistical summaries for multiple attributes at different granularities. To aid in this analysis, we worked in close collaboration with a transportation expert to design TR-EX, a visual exploration tool developed to identify, inspect and compare spatio-temporal patterns for planned and real transportation service. TR-EX combines two new visual encodings inspired by Marey's Train Schedule: Trips Explorer for trip-level analysis of frequency, deviation and speed; and Stops Explorer for station-level study of delay, wait time, reliability and performance deficiencies such as bunching. To tackle overplotting and to provide a robust representation for a large numbers of trips and stops at multiple scales, the system supports variable kernel bandwidths to achieve the level of detail required by users for different tasks. We justify our design decisions based on specific analysis needs of transportation analysts. We provide anecdotal evidence of the efficacy of TR-EX through a series of case studies that explore NYC subway service, which illustrate how TR-EX can be used to confirm hypotheses and derive new insights through visual exploration.

Index Terms—Transportation, schedules, kernel density estimation, visual exploration

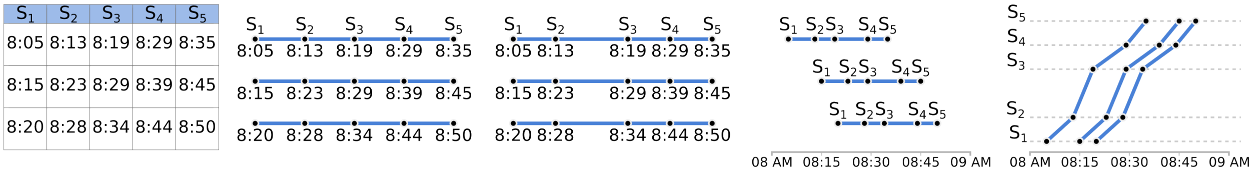
1 Introduction

Schedules play an important role in providing information to urban transit commuters for route and trip planning. Transportation agencies plan schedules to satisfy multiple criteria such as guarantee safe operation, efficient energy consumption, high scalability, and target levels of service efficiency which involve wait time and reliability measures. Inaccurate schedules not only negatively impact the users' experience but they also decrease service reliability. In practice, real service usually deviates from the planned schedule due to multiple reasons, including track or route sharing, ridership fluctuations, extreme weather, accidents and delays inherent to the dynamic behavior of transit systems. It is thus crucial for transportation analysts to understand the issues so that they can propose effective solutions that take into account installed capacity, demand, budgetary and personnel constraints. In this paper, we propose a visual analytics framework that helps analysts with this task by allowing them to explore planned and actual transit service.

We collaborated with a transportation expert who has conducted multiple studies on the quality of public transport using purely statistical tools to compute marginal distributions, extrema, mean and variation of accuracy, reliability, wait time and speed per station and over time. A reported difficulty in this process is the need to create multiple models with different granularities. For example, the comparison of adherence to timetable for different routes depends on a high-level overview of coarser data, whereas low-level inspection at specific time periods and stations potentially reveals ill behavior, but intrinsically demands finer-grained analyses. Additionally, these analyses are often performed in isolation, but in general the non-trivial relationship between properties of multiple schedules are the causes of structural problems whose solutions enable better timetable design.

C. Palomo, Z. Guo, C. Silva, and J. Freire are with New York University.

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467592



(a) Tabular (b) Logical (c) Spatial (d) Temporal (e) Graphical Schedule

Fig. 2: Different transit schedule representations. *Graphical Schedule* (a.k.a. *Marey's Graph* or *Train Schedule*) provides a concise representation of *Tabular* and *Logical* information (stops organized within trips), *Spatial* (distance between pairs of stations), and *Temporal* (absolute and relative stop times). The x axis represents time of day, while the y axis represents the stations along a route, spaced according to physical distances.

To address these challenges, we propose TR-EX (Figure 8). To the best of our knowledge, TR-EX is the first interactive visualization tool that supports the analysis of planned and real service data for specific tasks, from overview to detail, at trip and station-level. Our work was inspired by Marey's Train Schedule [40], a commonly-used visual representation for static schedule data that is familiar to transportation experts. As Figure 2e shows, each transit trip is represented as a polyline with vertices defined by stop times at stations for a given transit route. The horizontal axis represents time of day, and the vertical axis lists the sequence of stations in the vehicles trajectory. Although overwhelming at first glance, this compact but powerful representation communicates several properties, cleverly combining high-level and detailed information. Figure 2 shows how it combines multiple aspects of schedules into a single view: stops organized within trips (logical); order of and distance between stations in the vertical axis (spatial); and time of trip start and stops in the horizontal axis (temporal). This information allows users to infer the trip direction, speed (the slop of the lines), headway (interval between two stops for a given station), trip duration, and frequency over the day.

In contrast to previous work [4, 22, 30], TR-EX displays multiple attributes available in Marey's Graph for analyses for both planned (the planned stop times for a trip) and actual service (the observed stop times for that trip). Although familiar to transportation analysts, Marey's Graph is not effective for the analysis of real service data, which consists of tens of thousands of trips over several months. Besides creating new challenges for interactive exploration, these large data sets also lead to cluttering – too many lines need to be visualized, thus hindering the extraction of meaningful information. A key insight to deal with clutter is to draw an analogy between Marey's Graph and parallel coordinates [21]: each trip corresponds to a data point in parallel coordinates, and stations are data dimensions. This allows the application of techniques successfully used for parallel coordinates to solve cluttering in this domain. TR-EX uses kernel density estimation [34, 37], which enables the exploration of the full spectrum of detail required for different tasks, at interactive rates.

The main contributions of this paper can be summarized as follows:

- We define the problem of interactive visual analysis of transportation schedules and discuss the desiderata for a principled solution.
- We extend Marey's Graph to display a large number of both planned and actual trips, enabling the interactive identification of patterns, analysis of frequency, deviation and speed.
- We propose a new encoding for the analysis of indirect attributes at station-level: delay, wait time and reliability.
- We implemented these techniques and integrated them in TR-EX, a visual analytics system that supports the exploration of planned and actual schedules. Since user interaction plays a fundamental role in the system, we follow Shneiderman's mantra in our design: *overview first, zoom and filter, then details on demand* [38].
- We present case studies carried our by transportation experts that show the benefits and limitations of the tool and proposed visual representations.

The remainder of the paper is organized as follows. Related work is discussed in Section 2. In Section 3, we discuss that data and analysis requirements for transportation analysts that guided the design of the new visual representations we propose. The proposed design and implementation of visual encodings for analyses at trip- and station-level are presented in Section 4. In Section 5, we provide anecdotal evidence of the efficacy of TR-EX through case studies and evaluation carried out by transportation experts that applied the tool study NYC subway service. We conclude in Section 6 where we discuss the limitations of our approach and outline directions for future work.

2 Related work

Transportation and human mobility are important topics that have received increasing attention from the visualization community in recent years [1, 5]. In this paper, we take a first step towards addressing a new problem in this domain: the exploration of planned and actual transportation services.

Real-time transit data has been successfully used to provide riders with more accurate stop estimates. One example is *OneBusAway* [13], a set of tools that turn real-time data about the location of buses into more accurate arrival time estimates provided to commuters. Reported positive effects include increased satisfaction and decreased wait time, both real and perceived. Real-time information can be used not only to inform riders about the current service state, but it also drives the work of schedule designers. Transport agencies regularly update schedules to reflect service changes and infrastructural issues to improve adherence and satisfy wait time targets. Based on the planned schedule and on historical data, analysts use statistical tools to compute better time tables according to different metrics.

Automatic design of timetables is a complex optimization process that involves multiple components such as target trains running time, capacity allocation, buffer time to avoid delay propagation, and dwell time [15]. Planned timetables have been used to create graphs representing traffic flows and physical topology of railroads [27]. Our task at hand, however, is to visually explore existing schedules using data about the real service, and not simulate traffic or railway topology. Station-centric visualization approaches have been proposed that use data for planned schedules. The method proposed by Joyce [24] helps users identify the fastest routes from one station to others, using data for California's Caltrain. Krause et al. [25] propose a similar method for buses in Konstanz, Germany, with more elaborate graph layout techniques to display routes. The focus of both techniques however is on journey planning, making them suitable for riders but not for transit planners or for analyzing schedules. Ochiai et al. [32] proposed a framework for analysis and forecasting of train operation. Ushida et al. [41] proposed methods to visualize train trips to better understand delays in the system, and used this information to modify timetables and make them more robust. Unlike these approaches which make use of static visualizations to answer specific questions about train operation, our goal is to support interactive exploration of the information.

Train Schedule (aka Marey's Graph), designed by Ibry and published by Marey in the 1880s [40], displays both a high-level overview and details about trips and stops. Previous works have extended this representation in different ways. Mai et al. [30] added visual cues for delay and passenger load information, but they do not support interactivity or flexibility in analysis of different attributes. Hranac et al. [22] proposed the visualization of adherence differences with a band between the mean actual service and the planned, which is colored to convey passenger load information. Byon et al. [4] discretized Marey's Graph into a heatmap representation to study schedule adherence, focusing mainly on bunching. Their matrix, however, is limited to the analysis of a single property per cell, the score for schedule adherence.

TR-EX uses Marey's Graph as the basis for its initial visual design, but differently from previous work, it enables exploration of multiple parameters present in the representation, for both planned and actual trips. The large number of trips for historical service data introduces cluttering issues that are common to parallel coordinates displays [21]. Many of the cluttering reduction techniques proposed for parallel coordinates are also applicable to Marey's Graph [9, 21]. These include clustering in the data domain [14, 31], bundling [44, 17, 33], screen-space [2, 23] and density-based approaches [20]. To avoid cluttering while maintaining interactive rates for exploration, TR-EX applies kernel density estimation [37] with GPU acceleration for each visualized attribute [28]. User interface elements to edit the kernel's bandwidth give users flexibility to define the required level of detail for specific analyses.

3 Data and Tasks

In this section, we describe the schedule data and present required tasks for exploration, understanding and comparison of schedules attributes.

3.1 Transportation Schedules

Planned Service. Transportation agencies design schedules for existing routes in the system based on constraints such as installed capacity, personnel, and target service levels in response to demand. Each **route** $S = (\langle s_1, d_1\rangle, \langle s_2, d_2\rangle, ..., \langle s_n, d_n\rangle)$ is defined by the sequence of stations and their position d_i along the route trajectory (e.g., stations in NYC subway Line 1). A route is served by a set of **trips**, where each trip $T = (t_1, t_2, t_3, ..., t_n)$ is a data point with epochs t_i for **stop times** of a vehicle at station s_i. Trips are usually clustered for different directions, e.g., Northbound versus Southbound, and days of the week, e.g., weekdays versus weekends. The set of trips with *desired* stop times for each *route* constitutes the **planned service/schedule**.

Actual Service. Since fluctuations in demand, accidents, delays due to overcrowding or shared installed capacity are more a norm than an exception in transport systems, the **actual service** (the set of trips with *observed* stop times) often differs from the planned schedule. Data about *real service* can span several months. In a system such as the NYC subway, there can be tens of thousands of trips per transit route.

Derived attributes. To compare planned and actual service, it is important to analyze attributes such as: **deviation** (difference between a planned stop time and the observed stop time for a trip); **headway** (distance or time between two stops at a *station*); **average wait time** ($\frac{1}{2}$*headway* assuming uniform distribution for riders arrival at a station); **reliability** (measured by the stability of wait times in specific stations over time). These attributes can be computed in a preprocessing step prior to the analysis.

3.2 Desiderata

Schedule designers need to study *planned* and *actual* service along with the derived attributes to identify, understand and compare different perturbations in service quality. TR-EX was designed to support the following specific analysis tasks:

T1: Compare planned timetables against real service. To measure quality of service, it is crucial to compare the behavior of actual trips against the planned schedule. Trends in frequency and deviation are measures that capture accuracy.

T2: Characterize speed profile at different route segments. Speed profile reasoning is important to prevent accidents, might suggest the need for maintenance inspection at specific locations, and works as a proxy for passenger load measures at stations over time, an invaluable piece of information when the installed system does not provides ridership information automatically. Besides, a progressively slower vehicle might cause *bunching*, a serious disruption in transit service that needs to be avoided.

T3: Assess delay, wait time and reliability at the station level. Visualization of station-level delay allows inspection to determine stop time adjustments. Wait time, one of the most important service quality metric, directly impacts client satisfaction [8, 29]. Reliability is best captured as the variation in wait times: stable wait times (even when longer) are perceived more positively by riders.

T4: Study the interplay of different attributes. The understanding of complex systemic deficiencies depend on detailed visual comparisons and analyses of multiple attributes. For example, lower speeds occurring together with short wait times suggest bunching. These events occurring in isolation, however, might be characteristic of different phenomena such as overcrowding (that cause lower speeds) or higher frequency (during peak hours wait time is naturally lower).

4 TR-EX

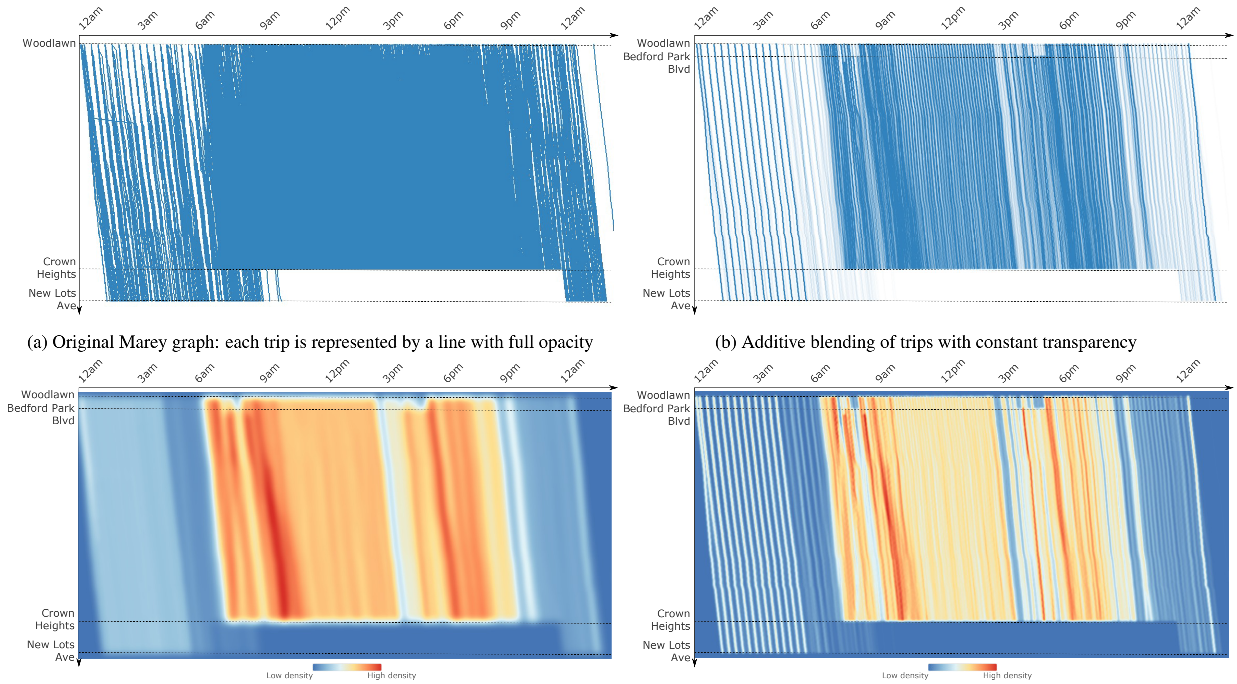
We now detail the design of TR-EX, which was guided by the tasks defined above. To illustrate the techniques, as a running example, we use the analysis of 10 weeks of NYC subway trips for *Line 4* – a total of 21,250 trips (654,252 stops) that occurred between October 6th and December 14th of 2014. In what follows, we motivate our use of Marey's Graph and discuss its limitations when used to explore actual trip data. We then present our approach to address these limitations. A contribution of our work is the adaptation of techniques from parallel coordinates to overcome cluttering. Next, we describe the two main visual components of TR-EX: *Trips Explorer*, for analyses of schedules at trip level; and *Stops Explorer*, for exploration of stops at station level. We conclude the section with a description of the different components of the TR-EX prototype and how users can interact with the system, including the use multiple linked views [36] with statistical summaries that support filtering of trips and stations.

4.1 Initial Design and Challenges

Since Marey's Graph is a representation that is familiar to transportation experts, in our first design, we used it to visualize the frequency of NYC subway trips. In Figure 3a, we show the visualization for downtown trip for subway line 4, where each trip is represented as a fully opaque blue line. We kept the convention of showing trips starting in a 24-hour time range, and displaying one line per trip. Patterns for ending stations are noticeable in Figure 3a. Trips seem to share the same starting station (Woodlawn, at the top), but there are two final stations: Crown Heights - Utica Ave for trips between 9am and 11pm, and New Lots Ave for trips after 11pm and before 9am. Notice that the high volume of data leads to considerable overplotting. Characteristic patterns such as frequency changes throughout the day are not visible. This can be alleviated with use of line opacity and additive blending to indicate density of the overlapping lines [43], as illustrated in Figure 3b. Higher-frequency patterns during peak hours (6-9am and around 6pm) and different starting stations (Woodlawn and Bedford Park Blvd) are then revealed.

This approach was used in the initial prototype of TR-EX, but several drawbacks were identified by expert users. Fixed transparency fails to convey details at higher frequency regions, such as service during peak hours. Experiments with transfer functions [23] to highlight specific features such as outliers or trends did not work well for analyses of different attributes.

In our discussions with schedule analysts, it became clear that their tasks consist primarily of identifying and investigating hotspots, regions with high or low densities. These hotspots are defined by *events* that deviate from normal behavior. For example, trips and stops that take longer than normal, or spatio-temporal regions where stops and trips present unusually high delays. Since analyses involve different



(a) Original Marey graph: each trip is represented by a line with full opacity

(b) Additive blending of trips with constant transparency

(c) Using KDE for lines: wide bandwidth reveals coarser details

(d) Using of KDE for lines: narrow bandwidth reveals finer details

Fig. 3: Marey's Graph applied to downtown trips of NYC subway line 4. Data about actual service consists of a large number of trips, leading to severe cluttering (a), even when additive blending (b) or transfer functions are used. The use of KDE avoids overplotting by revealing trends at different levels of detail: wide bandwidth shows a high-level overview (c), and a narrow bandwidth shows fine-grained details (d). In TR-EX, users can interactively select the bandwidth size according to their needs.

levels of detail, it is important to give users the ability to adjust the level of abstraction.

4.2 Reducing Clutter with Kernel Density Estimation

While aggregation and sampling can help alleviate overplotting effects, they do so at the cost of information loss [18]. Since TR-EX was designed for support interactive visualization of multiple properties of data sets consisting of a large number of trips, an effective solution must be 1) flexible, allowing exploration at different levels of detail, and 2) compute results at interactive rates, ideally without requiring pre-computation. These requirements make approaches such as Continuous Parallel Coordinates [20] and bundling [33] undesirable, since they either require pre-computation or prior knowledge of the data granularity that might reveal interesting patterns.

A compelling alternative is the use of Kernel Density Estimation (KDE), a non-parametric approach used to estimate probability density function of a random variable [37, 34]. KDE has been extensively applied in information visualization. Lampe and Hauser [28] proposed an interactive implementation of KDE that makes no assumptions about the data and uses the GPU to operate at interactive rates. The key idea is to perform accumulations of rendered elements in high-precision 2D textures, and later compute the KDE by rasterizing 2D polygons on the GPU. A strong benefit is that the time and space complexity of the KDE step does not depend on the data set size, since work is done in screen space.

As we discuss below, TR-EX uses KDE both for *Trips Explorer* and *Stops Explorer* as a flexible and unified approach to reveal trends at different scales. KDE calculates the probability density $\hat{f}$ at a given location l by weighing the attribute values of the spatial neighbors of l. Given a point x, its estimate will depend on the distance between x and every other data point x_i, weighted by a kernel function K, and a smoothing parameter (or bandwidth) h:

$$\hat{f}(\mathbf{x}) \quad = \quad \frac{1}{nh}\sum_{i=0}^{n} K(\frac{\mathbf{x}-\mathbf{x_i}}{h}) \tag{1}$$

The bandwidth determines the width of the kernel function; the kernel function determines the shape of the weighing function. K is usually a symmetric function that integrates to 1. It has been shown that the choice of the kernel is less important than an appropriate choice for bandwidth value [3, 39], so we restrict our discussion to the Gaussian kernel, shown in Equation 2:

$$K(x) \quad = \quad \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{2}$$

User-Defined Bandwidth. For continuous reconstruction of the distribution of an attribute c other than frequency, TR-EX uses the technique by Lampe and Hauser [28], which can be seen as the integral of a height field over the 2D domain that communicates the accumulated sum of all values c_i. We give users the control over the smoothing factor in KDE by providing UI controls for the bandwidth shape. As Figure 3 shows, different bandwidth sizes allow analysts to identify trends at different levels of detail. In Figure 3c, peak hours are highlighted in red and late-night trips are smoothed. In contrast, Figure 3d reveals additional details about late-night trips and trends during peak hours.

Estimation at Interactive Rates. Note that precise line density estimation [28] is too costly to achieve interactive rates for tens of thousands of lines, since for each line segment a quad needs to be rendered along which the density estimation is to be computed. This is not feasible for TR-EX, since the complexity is bound by the data set size. Consider our example data set for the NYC subway line 4: polygons would have to be rasterized for each of the 654,252 stops, making it very challenging (or even impossible) to obtain interactive rates. Instead, we first accumulate line frequencies or attributes in

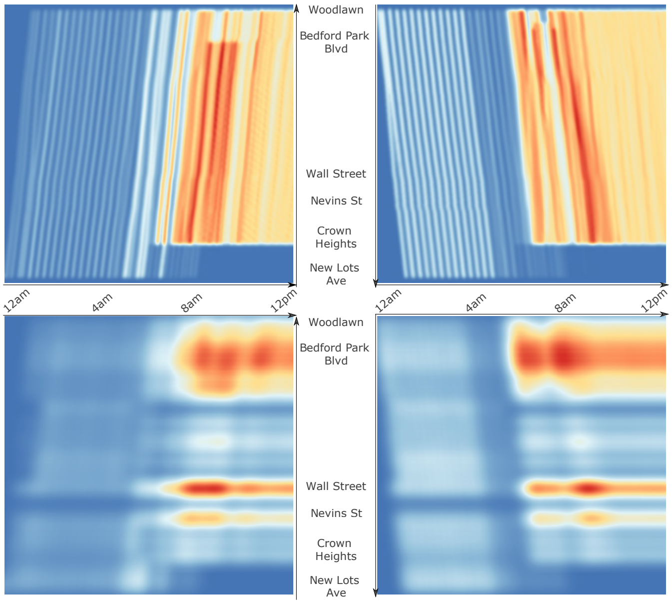


Fig. 4: Basic layout for Trips Explorer (top) and Stops Explorer (bottom): horizontal axis for time of day and vertical axis for stations along route. To avoid clutter, uptown (left) and downtown (right) trips are visualized separately, but stations are fixed for both directions of the route allowing users to go back and forth from one direction to the other without losing context.

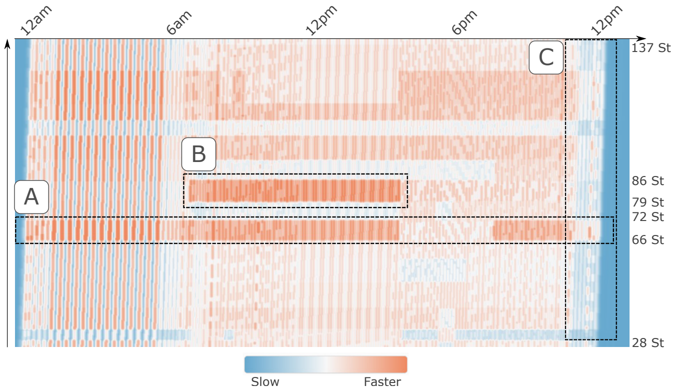


Fig. 5: Speed visualization with Trips Explorer for uptown trips in subway line 1. Region A shows that the speed for trips between 66th St and 72th St stations is mostly constant, except during peak hours, when vehicles run slower. Localized regions depict vehicles running faster than usual (region B) or slower during late nights (region C).

high-precision textures. Then, we later apply 2D KDE on the resulting texture, assuming the accumulated lines define a height field of frequency or an attribute such as wait time or delay, and the estimated density is done as for the point-wise density computation. This design choice allows TR-EX to achieve interactive rates and still gives users a good overview of when (x axis) and where (stations along y axis) interesting patterns occur.

4.3 Trips Explorer

The *Trips Explorer* extends Marey's Graph for use with a large number of real trips. As in the original design, we display stations along the vertical axis, spaced according to their physical distance along the route (see Figures 3c and 3d). One difference from the original design is that we limit the visualization of trips to a single direction: the simultaneous visualization of both directions causes cognitive overload and patterns are barely visible. Besides, we fix the order of stations in the vertical axis for both directions: when visualizing south-bound trips, trips run from top-left corner (north-most station, earliest hour) to bottom-right corner (south-most station, latest hour of the day) of the display. North-bound trips run from bottom-left corner to top-right corner. As shown in Figure 4, this choice helps users keep a mental map of the physical position (and orientation) of the stops when switching back and forth between directions, a common practice during exploration.

User interface (UI) elements and interactive summary plots provide users with the ability to choose to visualize planned or real service, different routes and directions, and filter trips for different days of the week. Color scales are also provided to convey measures of the different attributes available: users can select multiple options for qualitative and sequential scales from ColorBrewer [6]. Users can compare real and planned service in terms of frequency (through accumulation of lines density and then KDE computation as discussed in Section 4.2) and scalar attributes of trip such as deviation and speed.

Deviation. Trip deviation is defined as the distance measure between a polyline for a real trip and its matching planned trip, which is precomputed for each planned/real trip pair. We use L_1 distance to measure deviation of real stop times against planned stops. This measure highlights regions in time where the schedule suffers from greater perturbation, which negatively impacts the perceived service by riders. It also shows adherence: if deviations are frequent for all times of day, the planned schedule could be shifted to better represent the actual service.

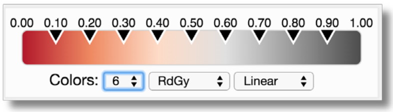
TR-EX allows direct visualization of deviation through sequential color scales, with all trip segments in a real trip (polyline) carrying the deviation distance. As explained in Section 4.2, after the initial step of accumulation of that attribute, the resulting density estimation of the KDE step is mapped to a chosen sequential color scale (continuous or discrete) defined by the user.

Speed. Analysts use information about trip speed to communicate possible structural problems to maintenance departments, and to assess whether safe speed limits are followed by conductors. The *Trips Explorer* encodes vehicle speed in the slope of the lines: steeper lines represent trips segments with higher speed. However, the comparison of angles requires fine inspection and is impractical for higher zoom levels, where slopes are barely visible. TR-EX also applies the same technique used for displaying deviations to the visualization of speed, with speeds set per trip segment. With the scalar value for speed per trip segment, TR-EX also provides filters by higher or lower speeds to improve exploration of specific speed ranges. The proposed approach for speed visualization displays regions of slower and faster trips, without forcing the users to closely inspect the angles of the lines as in Marey's original design. As Figure 5 shows, this representation clearly shows track segments with uniform speed vehicles throughout the day (region A), with localized faster (region B) or slower (region C) vehicles during specific time ranges. Notice that speed and deviation are not interpolated between pairs of vertices, instead, they defined per trip segment (adjacent pairs of stations).

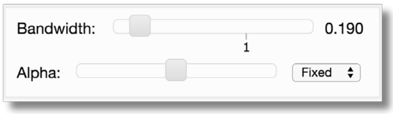
Interaction. To allow inspection of specific times of day, we provide zooming and panning on the plot (horizontal axis). Zooming in the vertical axis can be done through brushing of specific stations on the route (see component 4 in Figure 8). A color editor (Figure 6a) lets users define different color scales (all from ColorBrewer [6]), the number of colors, choose between discrete and linearly interpolated colors, and edit the color stops to change the mapping of scalar values to specific ranges. Also editable are the opacity multiplier for scalar values and the KDE smoothing parameter defined by the bandwidth, which can be seen in Figure 6b.

4.4 Stops Explorer

The *Trips Explorer* allows users to analyze trip behavior. However, some schedule attributes can only be analyzed at the station level, notably: delay, wait time and reliability. Each of these attributes is a scalar value associated with a vertex (the stop). TR-EX exposes this information by incorporating an additional visual representation, the *Stops Explorer*, shown in Figure 7). Similar to a dot plot or Cloud-



(a) Editable color scale



(b) KDE bandwidth and alpha multiplier controls

Fig. 6: UI interaction elements to customize the rendering of trips and stops give users flexibility during specific analyses. Mapping of scalar values is controlled through an editable color scale, while the KDE bandwidth can be adjusted to reveal more or less detail.

lines [26], it shows planned and actual trips in detail for each station, but keeps the same axes conventions used in the *Trips Explorer*: time in the horizontal axis and stations distances in the vertical axis. This choice keeps the representation uniform across visualizations, allowing analysts to switch back and forth between trip and station analyses in a seamless fashion.

A design analogous to *Trips Explorer* is used: each stop is drawn as a dot at the station level carrying the visualized scalar attribute, and after accumulation, the result of the KDE step is visually mapped with a sequential color scale defined by the user. Multiple coordinated plots for the different attributes are available for filtering and further investigation. Figure 7 shows stop delays for trips on subway line 4 occurring early in the day on weekdays. Notice the shared delay patterns for downtown (left) and uptown (right) trips. At the 125 St Station for both downtown (A) and uptown (B) trips, delays start to increase. This station is the last one uptown Manhattan, so most Bronx residents commuting back from Manhattan might take the train at that station, while the opposite is also true for residents of lower Manhattan and Brooklyn leaving upper Manhattan taking southbound commute. Higher demand results in longer boarding times and subsequent increase in delay. In Section 5, we discuss case studies for the two other available attributes at station-level, namely wait time and reliability.

4.5 Putting It All Together: The TR-EX Prototype

Project History. In 2012, members of our team met with analysts from the New York Metropolitan Transportation Authority (MTA) to discuss potential projects. At that time, the MTA was implementing a real-time system to make train and bus times available using the General Transit Feed Specification (GTFS) [16]. We started to collect the subway GTFS feed when it first became available, and spearheaded an effort to create analysis tools for the data. This effort was the beginning of the development of TR-EX.

During 2013, we experimented with several designs. We first tried to use geographical metro maps to display the location of the trains. Although visually engaging, we found it hard to use as an analysis tool. We got naturally drawn to Marey's Graph. Not only do they concisely and effectively convey different aspects of schedules, but also transportation experts are very familiar with the representation they use. A key focus of our work was to extend Marey's Graph to support interactive exploration of a large number of trips. In early 2014, we had the opportunity to meet with MTA analysts to obtain feedback on our work. The tasks listed in Section 3.2 come to a large extent from those discussions and our own findings during 2013.

System Overview. An overview of TR-EX interface is shown in Figure 8. The system enables transportation analysts to carry out the analysis tasks described in Section 3.2. The comparison between planned and actual service (*T1*) can be done through a combination of *Trips Explorer* and *Stops Explorer*. The *Trips Explorer* conveys information about trip speeds and allows the identification of misconduct or maintenance needs, as required by task *T2*. The *Stops Explorer* was

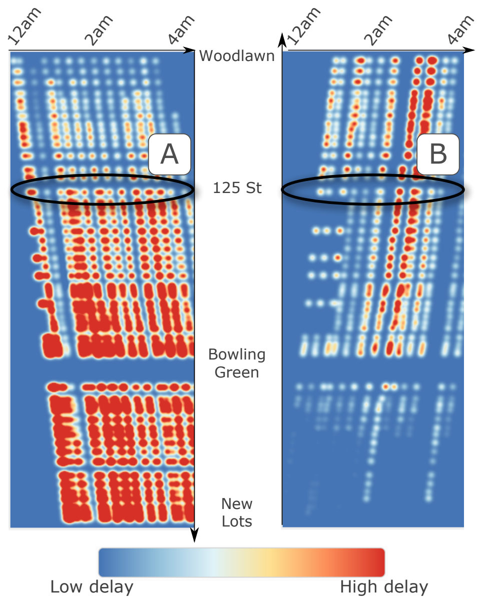


Fig. 7: Stops Explorer: This visualization shows the delay at specific stations for subway line 4, for trips that take place at early hours of the day on weekdays. Regions A and B show shared behavior for uptown and downtown trips, with increase in delay at (or right after) that station, likely due to higher demand and slower boarding times.

specifically designed to address the requirements of task *T3*. With Trips Explorer and Stops Explorer, analysts can investigate schedule properties for actual and planned service, at trip- or station-level, and compare different attributes as required by task *T4*.

For fast prototyping and greater convenience to collaborate with experts, TR-EX was implemented as a browser-based application. A light-weight web-server stores the preprocessed data set of real and planned service. To allow interactive visualization of thousands of trips, rendering is done through WebGL and HTML5 Canvas. Components used as summaries and filtering tools were developed with client-side libraries such as D3.js and jQuery.

Rendering. The volume of data required us to think carefully about how to implement the underlying renderer. After initial data set loading from the server, geometry is created for planned and real trips and stops, and sent to the graphics card at startup. Then, all UI elements that change the behavior of *Trips Explorer* and *Stops Explorer* are mapped to `uniforms` (which are variables that have the same value for all vertices) for range of attributes visualized, view matrix in response to zoom or pan and attribute visualized. The only exception is the editable color scale used to map KDE results into color, which is stored as a texture in the GPU and is updated when users either change the color scheme, number of colors or color scale stops. With this approach, TR-EX allows interactive visualization of tens of thousands of trips (or millions of stops).

User Interface. The TR-EX user interface shown in Figure 8 was designed to be modular with the idea that it would be customizable for particular end-user applications. The interface shown in the figure has a number of visualization and selection tools, which are described in detail below.

The Data Selection Panel (box "1") allows users to choose a subway line to explore, and a particular subset of the trips in that line, including Northbound vs. Southbound traffic, and Planned vs. Real arrival times. It would be natural to also include time ranges there, although we have decided for now to allow direct manipulation of time ranges by dragging and zooming in the Main Canvas (box "3").

The selection between Trips Explorer and Stops Explorer is performed by selecting the appropriate tool (box "2"). The attributes to be displayed are selected in box "6". The Filter Panel (box "5") is used

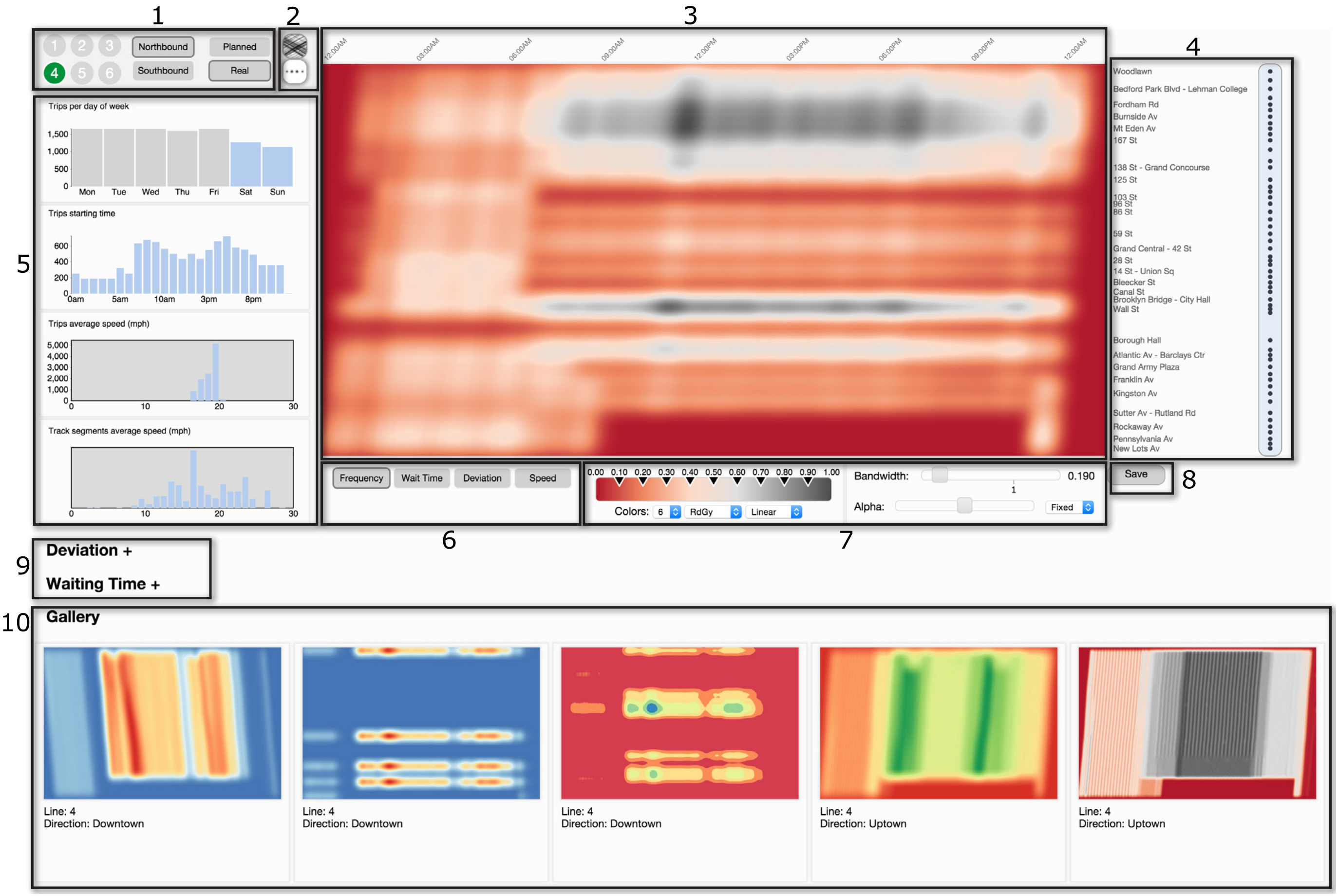


Fig. 8: Overview of the TR-EX proof-of-concept prototype. Routes, directions and real/planned service can be selected in (1). Users can choose between *Trips Explorer* and *Stops Explorer* (2) for the main plot (3). The Subway Station list (4) shows the stations in the line and allows users to filter through brushing. The filters panel (5) provides controls for setting visibility for specific days of the week and filtering of speed ranges for trips and segments. TR-EX allows users to perform analyses based on different service attributes (6), in this case, frequency of stops is selected for line 4 uptown trips on weekends. Users can also adjust the bandwidth size and the mapping of KDE result to specific color scales (7). TR-EX provides additional summary plots for specific attributes (9), and users can save the current visualization state (8) to a gallery (10) for further comparison.

both for data display and selection through the use of brushing and linking. TR-EX displays the relative data distribution through graphs that can be selected, e.g., in Figure 8 we are selecting the data for Saturday and Sunday (they are the two blue bars in the first plot of box "5") to be displayed in box labeled "3", which is the main rendering canvas. The list of Subway Stations (box "4") can be used for selection through brushing. The save button in box "8" is used to save a particular configuration to the Gallery, shown in box "10". Additional visualizations with brushing and linking functionality are available on demand (box "9"). Since this functionality is not relevant to this paper, we refrain from describing all the other plots.

Values for the kernel size and type used for Kernel Density Estimation can be set in box "7". As we mentioned before, the results reported in this paper, a gaussian kernel was used with a symmetric axis-aligned bandwidth, i.e., KDE was computed with bandwidth set by the user over a fixed-size 2D 64×64 pixels patch in the screen-space.

5 System Assessment and Feedback

We report on two different assessements we carried out for TR-EX. First, we describe the structured collaboration that guided our design of the TR-EX prototype and functionality, and case studies that highlight the effectiveness of the system in helping analysts explore wait time and reliability. Second, we present an interview with a transportation expert with over 20 years of experience that works with MTA on addressing real-world problems. Last, but not least, we show how TR-EX can help experts identify problems in the data.

5.1 Case Studies: Studying Wait Time and Reliability

We collaborated with a transportation expert on the design of TR-EX. Over a period of six months, we had weekly meetings. Feedback and suggestions gathered in these meetings contributed to the design and evolution of the TR-EX prototype. In his words, "With this tool (TR-EX), I have access to the full spectrum of detail I need for my analyses, from high-level understanding of general behavior to fine inspection of specific service deficiencies such as bunching". In this section, we present anecdotal evidence of the efficacy of our proposal through case studies carried out by our collaborator, where he used TR-EX to analyze NYC subway trips for lines 1-6 occurring between October 6th and December 14th 2014.

5.1.1 Exploring Wait Time

Wait time is one of the most important service quality metrics, since it directly impacts riders' satisfaction [8, 29]. Wait time is computed as half headway, assuming a uniform distribution of passenger arrivals at stations for high-frequency transit services. Similar to other schedule attributes, real wait time (computed with headways between observed stops) might differ from the planned wait time (computed with headways between planned stop times).

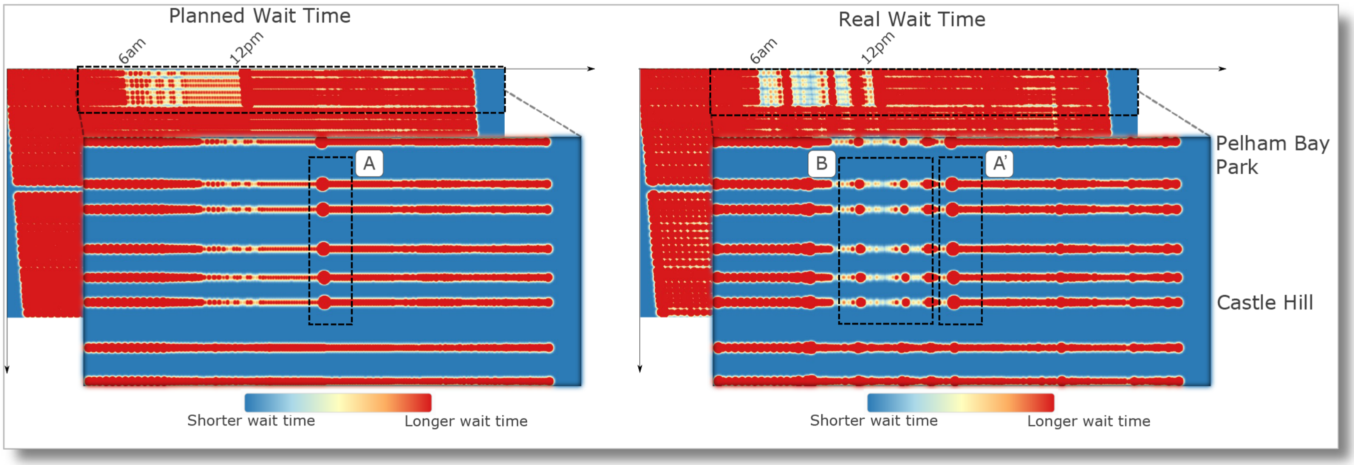


Fig. 9: Comparing planned (left) and observed (right) wait times at stations for subway line 6 toward Brooklyn Bridge - City Hall on weekdays. The visualizations show a substantial divergence between the planned and actual wait times in the highlighted region (top). Zooming into this region, we can see that in the planned service, stops are equally spaced from each other, while in the real service trips are concentrated in three main clusters: time: around 8am, 10am and 11am. This leads to a considerable increase in observed wait time at later times, as shown in region *B* of the real service. The behavior of the real service stabilizes after region *A'*, where it is similar to the planned schedule.

One example of wait time analysis is shown in Figure 9, for subway line 6 trips downtown, toward Brooklyn Bridge - City Hall, which shows planned (left) and actual (right). A prominent pattern is detected between 6am and 12am, from Pelham Bay Park to Castle Hill: the observed wait times differ considerably from the planned ones. While in the planned service stops are equally spaced from each other in time, the real service shows a concentration of trips in three clusters: around 8am, 10am and 11am. Observed wait times increase substantially later. The ability to easily identify wait times issues such as this one helps transportation experts focus on trying to understand why such patterns arise and on strategies to address the problems, which often spur the need for additional explorations. TR-EX therefore shows great potential to streamline the exploration, analysis and refinement process.

5.1.2 Assessing Reliability

For high frequency urban rail like the MTA subway, reliability is best captured by the fluctuation of wait time experienced by passengers at the same time period at a particular station. Therefore, reliability can only be accurately measured when all data points, and not a sample, is observed. This is only feasible when the operation data are collected electronically through automatic vehicle location (AVL), automatic passenger count (APC), smart cards, etc.

We capture the dispersion of wait times with the coefficient of variation (CV) : $CV = standard\ deviation / mean\ actual\ wait\ time$ (or $\frac{\sigma_{wt}}{\mu_{wt}}$). CV normalizes the differences in wait time by time periods, e.g., peak hours have shorter wait times than off-peak hours. The larger the CV, the greater the variation in the wait time, and lower the reliability, which is defined as:

$$reliability = 1 - \frac{\sigma_{wt}}{\mu_{wt}} \tag{3}$$

The analysis of reliability of subway line 1 weekdays trips toward Van Cortlandt Park 242nd Street is depicted in Figure 10. The expert observed interesting patterns, including some that were previously unknown. Stations with lower demand often provide higher reliability, which is confirmed with the visualization for Christopher Street station (region *B*), which presents high reliability throughout the day. Another hypothesis confirmed by the expert was that the subway is less reliable during peak hours: with higher frequency service, stop times vary considerably, and so do wait times. That fact can be verified for the 23rd St station during evening peak hours (region *C*). An unexpected pattern, however, was observed between 14 Street Station and 28th St, after morning peaks and before 2pm (region *A*). All stations in the segment (which covers the Chelsea neighborhood) have high demand throughout the day, but the expert was intrigued by the localized low reliability behavior during those specific times, and will gather more information about boarding levels to further investigate this and try to understand potential causes.

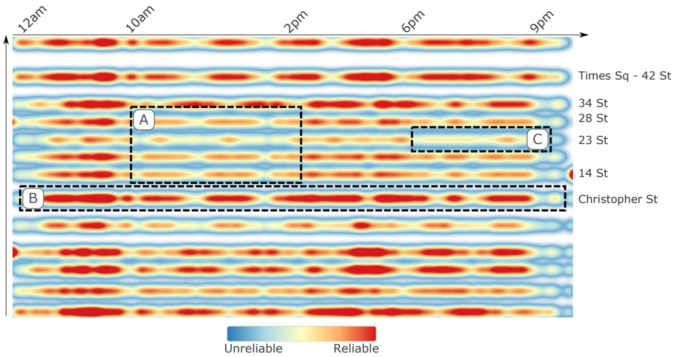
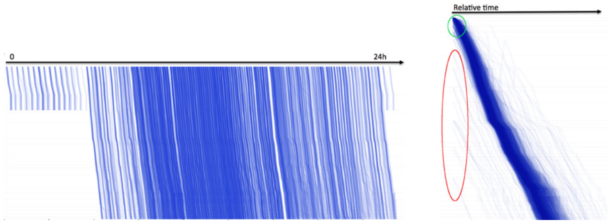


Fig. 10: Reliability visualization of weekdays trips for subway line 1 toward Van Cortlandt Park 242nd Street. Region *B* confirms the hypothesis that stations with lower demand are more reliable, and region *C* shows that peak hours cause considerable perturbations in wait time due to higher frequency of vehicles, and resulting in lower reliability. Region *A* in Chelsea presents unexpected low reliability between 10am and hours.

5.2 Interview with Transportation Expert

To validate our design decisions, we interviewed a transportation expert. Prior to the interview, he was given access to the system to explore its functionality. As we started the interview, he noted that “the visual metaphor was obvious”. This reinforced our rationale for selecting Marey’s Graph as the basis for TR-EX: the representation is familiar and easy to understand for transportation experts. The use of KDEs was less obvious and it took him a little while to understand how the process of going from “blurry” to “sharp” images worked. After this learning period, he liked the access to KDEs, since it allowed him to look at the data at different levels of scale, which displayed different features of the data, and could be “zoomed in” on demand.



(a) Trips starting in the 24h period (b) Translated

Fig. 11: Our visualization can also help in the data cleaning step, making it clear the structure of the real service dataset.

This multi-scale approach to visualization has been previously studied by Wattenberg and Fisher [42], where they highlighted the benefits of having the perceptual organization reflect the organization of the underlying data. He was also impressed with the ability to look at the different derived variables, including wait time and deviation, and that it was possible to compare them over a large set of trips.

After saying that TR-EX "looks great and it is an excellent tool", he gave us several suggestions for improvements. Most of his comments are orthogonal to the functionality already provided, but deals with features that would allow the system to be used for "actionable" analysis from an "end-user perspective". As he has been working closely with the MTA, the target users he has in mind are MTA staff members. He suggested that we augment our annotation capabilities to make it easier to share analysis; there is interesting work on these lines by Heer et al [19]. Another component he would like to see integrated into the system is a map view. This issue needs further exploration since we tried this early in the project and we were not satisfied with the results. He also suggested a number of features that could be added, including the ability to order stations based on their statistics, for instance, the stations that experience the most delays. Finally, he mentioned that we should explore using the system for bus schedules after suitable customization.

5.3 Identifying Problems in the Data

Building a data set with historical trips that is suitable for analysis is a non-trivial task. For NYC subway, for instance, to build the real-service dataset, we need to use real-time data provided by the MTA. The systems used to capture control signals and train movement can fail, and as result, real-time feeds often contain erroneous for trips, such as wrong or missing stop for stations. Matching actual and planned trips is also challenging, since for some control systems the association is not explicit and needs to be inferred using trips' start times, train identifications and headsigns.

TR-EX can assist in this initial process of data wrangling and cleaning. For example, Figure 11b shows Southbound trips for subway Line 3 that should all start at the Harlem - 147th St Station (circle in green), but many trips (inside red ellipse) actually have a different start station, making clear those trips are missing partial data. Notice how without the data transformation (Figure 11a) it is impossible to identify trips with missing data for multiple stations. With this easy identification, schedule analysts can take action to handle the unexpected trips format. Options include removing them entirely from the dataset, or set default values to complete the missing stops.

6 Conclusion

The analysis and visualization of transportation data is an area that has attracted considerable attention in the recent literature (see, e.g., Andrienko et al. [1] and Chen et al. [5] for comprehensive surveys on existing techniques), and where the visual analytics community can have substantial impact. In this paper we presented TR-EX, an interactive visualization tool designed to support the exploration of transportation schedules and service. This work is part of a larger research effort centered around the development of methods and tools to analyze complex transportation systems and human mobility in cities [10, 35, 7, 12]. Our work is grounded on real problems and user needs. The idea to develop TR-EX came from an ongoing with the MTA, the agency that runs the largest transportation system in the United States, and the project was carried out in a close collaboration between a team consisting of computer scientists and a transportation expert.

The visual representations used in TR-EX were inspired by Marey's Graph [40]. The *Trips Explorer* extends Marey's Graph for analysis of attributes at trip-level such as frequency, deviation and speed; and the *Stops Explorer*, an abstraction similar to dot plots, enables visualization of station-level attributes such as delay, wait time and reliability. For both representations, TR-EX uses 2D kernel density estimation in screen-space to make analyses of large datasets feasible and clutter-free. User-configurable bandwidth size solves overplotting while keeping the aggregation to the level required by the task at hand. Our assessment strategy included a number of case studies carried out by our key collaborators; interviews with domain experts; and our own assessment of the tool as visualization experts. We anticipate the usefulness of TR-EX will only increase given the growing volume transit data that are becoming available.

Our existing system has a number of limitations that we plan to address in future work. First, there are several shortcomings that come from pursuing a web-based approach. Although the web visualization technology has advanced considerably, it is harder to get consistent UI interaction and design for different browsers. The graphics technology is also behind that available for desktop applications. All of these challenges made development harder for us, but we believe it provides a stronger foundation for the future. Second, KDE provides a clear advantage for dealing with over-cluttering, but its use needs to be simplified if we want to reach an end user that lacks high-level visualization literacy. We also need to implement more sophisticated caching schemes that allows us to handle much larger datasets. Such datasets are likely to benefit from further improvements in the user interaction. Currently, users need to look at two different visualizations placed side by side to identify the differences between planned and actual wait times. A better solution would be to compute and display the differences directly, similar to the approach used to compare bird migration patterns proposed in [11].

There are many avenues for future research. One natural direction is to adapt TR-EX to support different means of transportation. In particular, buses, where bunching is a critical issue, and low-frequency systems such as trains, where adherence and time at stations become important aspects to visualize. TR-EX could also be extended with tools to compare different routes: currently each route's schedule is handled in isolation, but often, vehicles share the installed capacity, so issues in one route are likely to affect others. Another direction we would like to explore is the simulation of changes and the analyses of their impact on the service, for example, rerouting, inclusion or exclusion of vehicles, or service disruption at specific stations and times, and even try to model how different transportation models and disruptions might affect another.

Acknowledgments

We thank our anonymous reviewers for their thoughtful suggestions which helped us greatly improve this manuscript. We would also like to thank Professor Kaan Ozbay for providing us with detailed input on TR-EX. Also, we thank Ernest Tollerson, Dan Queally, and their colleagues from the MTA for all the help with GTFS access, and many discussions that helped to guide the development of TR-EX and shape our understanding of mass transportation. This work was supported in part by a Google Faculty Award, an IBM Faculty Award, the Moore-Sloan Data Science Environment at NYU, the NYU School of Engineering, the NYU Center for Urban Science and Progress, AT&T, and NSF award CNS-1229185.

REFERENCES

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

[2] A. O. Artero. Uncovering clusters in crowded parallel coordinates visualizations. In *Proceedings of IEEE Information Visualization 2004*, pages 81–88, 2004.

[3] J. Bithell. An application of density estimation to geographical epidemiology. *Statistics in medicine*, 9(6):691–701, 1990.

[4] Y.-J. Byon, C. E. Cortes, F. J. Martinez, M. A. Munizaga, and M. Zuniga. Transit performance monitoring and analysis with massive gps bus probes of transantiago in santiago, chile: Emphasis on development of indices for bunching and schedule adherence. In *Transportation Research Board 90th Annual Meeting*. Transportation Research Board, 2011.

[5] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 2015.

[6] Colorbrewer. `http://colorbrewer2.com`, 2015.

[7] H. Doraiswamy, N. Ferreira, T. Damoulas, J. Freire, and C. Silva. Using topological analysis to support event-guided exploration in urban data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2634–2643, 2014.

[8] A. Durrande-Moreau. Waiting for service: ten years of empirical research. *International Journal of Service Industry Management*, 10(2):171–194, 1999.

[9] G. Ellis and A. J. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.

[10] N. Ferreira, J. T. Klosowski, C. E. Scheidegger, and C. T. Silva. Vector field *k*-means: Clustering trajectories by fitting multiple vector fields. *Computer Graphics Forum*, 32(3):201–210, 2013.

[11] N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. Silva. Birdvis: Visualizing and understanding bird populations. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2374–2383, 2011.

[12] 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, 2013.

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

[14] Y.-H. Fua, M. Ward, and E. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of IEEE Visualization*, pages 43–508, 1999.

[15] R. Goverde, I. Netherlands Research School for Transport, and Logistics. *Punctuality of Railway Operations and Timetable Stability Analysis*. TRAIL thesis series. Netherlands TRAIL Research School, 2005.

[16] GTFS. `https://developers.google.com/transit/gtfs-realtime`, 2012.

[17] P. Guo, H. Xiao, Z. Wang, and X. Yuan. Interactive local clustering operations for high dimensional data in parallel coordinates. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 97–104, 2010.

[18] M. Hao, U. Dayal, D. Keim, and T. Schreck. Multi-resolution techniques for visual exploration of large time-series data. In *Proceedings of the 9th Joint Eurographics/IEEE VGTC Conference on Visualization*, pages 27–34, 2007.

[19] J. Heer, J. D. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, 2008.

[20] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1531–1538, 2009.

[21] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. In *STAR Proceedings of Eurographics 2013*, pages 95–116, 2013.

[22] R. Hranac, J. Kwon, M. Bachmann, and K. Petty. Using marey graphs to visualize transit loading and schedule adherence. In *Transportation Research Board 90th Annual Meeting*. Transportation Research Board, 2011.

[23] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Proceedings of IEEE Information Visualization 2005*, pages 125–132, 2005.

[24] A. Joyce. Automated visualization of public transportation time schedules. `http://thisisalan.com/other/joyce_transvis.pdf`, 2011.

[25] J. Krause, M. Spicker, L. Worteler, M. S. 0001, L. Zhang, and H. Strobelt. Interactive visualization for real-time public transport journey planning. In *SIGRAD*, volume 81, pages 95–98, 2012.

[26] M. Krstajic, E. Bertini, and D. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432–2439, 2011.

[27] M. Kurant and P. Thiran. Extraction and analysis of traffic and topologies of transportation networks. *Phys. Rev. E*, 74:036114, 2006.

[28] O. Lampe and H. Hauser. Interactive visualization of streaming data with kernel density estimation. In *Proceedings of IEEE Pacific Visualization 2011*, pages 171–178, 2011.

[29] Y.-W. Li. Evaluating the urban commute experience: A time perception approach. *Journal of Public Transportation*, 6(4):41–67, 2003.

[30] E. Mai, M. Backman, and R. Hranac. Visualizing bus schedule adherence and passenger load through marey graphs. *ITS World Congress*, 2011.

[31] M. Novotny and H. Hauser. Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900, 2006.

[32] Y. Ochiai, J. Nishimura, and N. Tomii. Punctuality analysis by the microscopic simulation and visualization of web-based train information system data. In *Computers in Railways XIV*. WIT Press, 2014.

[33] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. Seidel, and T. Weinkauf. An edge-bundling layout for interactive parallel coordinates. In *Proceedings of IEEE Pacific Visualization 2014*, pages 57–64, 2014.

[34] E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[35] J. Poco, H. Doraiswamy, H. T. Vo, J. L. D. Comba, J. Freire, and C. T. Silva. Exploring Traffic Dynamics in Urban Environments Using Vector-Valued Functions. *Computer Graphics Forum*, 34(3):161–170, 2015.

[36] J. 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, 2007*, pages 61–71, 2007.

[37] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

[38] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[39] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[40] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 1986.

[41] K. Ushida, S. Makino, and N. Tomii. Increasing robustness of dense timetables by visualization of train traffic record dataand monte carlo simulation. In *World Congress on Railway Research*, 2011.

[42] M. Wattenberg and D. Fisher. A model of multi-scale perceptual organization in information graphics. In *Proceedings of IEEE Information Visualization 2003*, pages 23–30, 2003.

[43] E. J. Wegman and Q. Luo. High dimensional clustering using parallel coordinates and the grand tour. *Computing Science and Statistics*, 28:361–368, 1996.

[44] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. *Computer Graphics Forum*, 27(3):1047–1054, 2008.