Visually Exploring Transportation Schedules

Paper By:Cesar Palomo, Zhan Guo, and Juliana

Presentation By:Huaying Tian

Outline

Framework → Previous Work-Marey's Graph → Methods-TR-EX → Comments/ Critiques

Framework

Fig. 1:What-Why-How

Domain : Transportation and Human Mobility

Target User: Transportation Analysts

Data:10 weeks of NYC subway trips for line 4-a total of 21,250trips (654,652 stops) that occurred between 6th October and 14th December of 2014.

Tasks:
T1:Compare planned timetables against real service
T2:Characterize speed profile at different route segments.
T3:Assess delay, waittime and reliability at the station level.
T4:Study the interplay of different attributes.

How:Encode sequential color scale; scalar value associated with a vertex (the stop);bar charts;dot plot;cloudlines; polyline with vertices

How:Reduce Filtering

How:Manipulate Navigate:Zoom and Pan

Transportation Schedules

1.Planned Service: The set of trips with desired stop times for each route constitutes the planned service/schedule.

2.Actual Service： The actual service (the set of trips with observed stop times) often differs from the planned schedule.

3.Derived Attributes： 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** (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.

Previous Work-Marey's Graph or Train Schedule

Although familiar to transportation analysts,not effective for the analysis of real service data, which consists of tens of thousands of trips over several months.

Marey's Train Schedule-a commonly-used visual representation for static schedule data that is familiar to transportation experts.

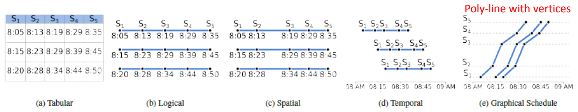


Fig. 2: Different transit schedule representations. Graphical Schedule (a.k.a. Marey's Graph or Train Schedule) provides a concise representation of Tabular and Logical information (stops organized within trips), Spatial (distance between pairs of stations), and Temporal (absolute and relative stop times).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

This paper takes a first step towards addressing a new problem in this domain: the exploration of planned and actual transportation services.

Pros and Cons of Marey's Graph

Pros:
1.Concisely and effectively convey different aspects of schedules

2.Transportation experts are very familiar with it

Cons:
1.Not effective for the analysis of real service data,which consists of tens of thousands of trips over several months;

2.Create new challenges for interactive action;

3.These large data sets also lead to cluttering – too many lines need to be visualized,thus hindering the extraction of meaningful information

TR-EX is proposed

Methods:TR-EX(Composed of Trips Explorer and Stops Explorer)

TR-EX is a visual analytics system for detection, inspection and comparison of spatio-temporal patterns in transportation services.

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

An Overview of TR-EX System

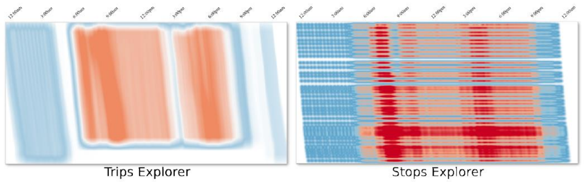


Fig. 3: TR-EX is a visual analytics system for detection, inspection and comparison of spatio-temporal patterns in transportation services. It uses and combines new visual encodings inspired by Marey's Graph to support the exploration of planned and real transportation schedules, showing where and when 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.

Strong benefit: The time and space complexity of the KDE step does not depend on the data set size, since work is done in screen space.

Reducing Clutter with Kernel Density Estimation (KDE)

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. 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 .

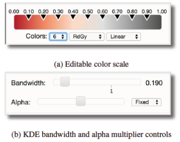
KDE calculates the probability density ˆ *f* at a given location *l* by weighing the attribute values of the spatial neighbors of *l*. Given a point *x*, its estimate will depend on the distance between *x* and every other data point *xi*, weighted by a kernel function *K*, and a smoothing parameter (or bandwidth) *h*:

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

The bandwidth determines the width of the kernel function; the ker-nel 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 , so we restrict our discussion to the Gaussian kernel, shown as follows:

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

User Interaction of Trips Explorer



give users the control over the smoothing factor in KDE by providing UI controls for the bandwidth shape

Fig. 4: 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.

Marey's Graph

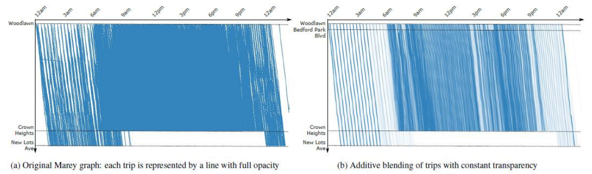


Fig. 5: Marey's Graph applied to downtown trips of NYC subway line 4. Data about actual service consists of a large number of trips, leading to severe cluttering (a), even when additive blending (b) or transfer functions are used.

Characteristic patterns such as frequency changes throughout the day are not visible!

With KDE

User-Defined Bandwidth!

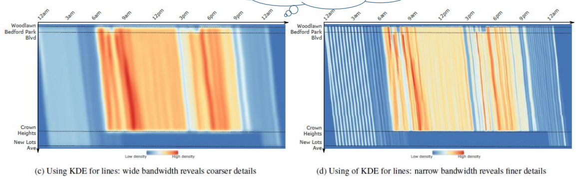


Fig. 5: Marey's Graph applied to downtown trips of NYC subway line 4. Data about actual service consists of a large number of trips. The use of KDE avoids overplotting by revealing trends at different levels of detail: wide bandwidth shows a high-level overview (c), and a narrow bandwidth shows fine-grained details (d). In TR-EX, users can interactively select the bandwidth size according to their needs.

Give users the control over the smoothing factor in KDE by providing UI controls for the bandwidth shape. As Figure 5 shows, different bandwidth sizes allow analysts to identify trends at different levels of detail.

TR-EX:

Trips Explorer：Extends Marey's Graph for use with a large number of real trips,allows users to analyze trip behavior.

Stops Explorer:At the station-level,some attributes like delay ,wait time and reliability can only be analyzed at that level.It shows planned and actual trips in detail for each station,but keeps the same axes conventions used in the Trips Explorer. This choice keeps the representation uniform across visualizations, allowing analysts to switch back and forth between trip and station analyses in a seamless fashion.

Trips Explorer and Stops Explorer

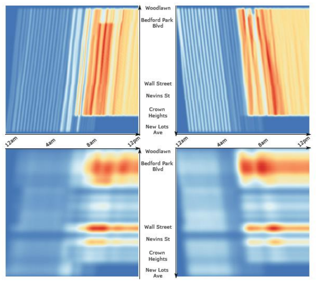


Fig. 6: Basic layout for Trips Explorer (top) and Stops Explorer (bottom): horizontal axis for time of day and vertical axis for stations along route. To avoid clutter, uptown (left) and downtown (right) trips are visualized separately, but stations are fixed for both directions of the route allowing users to go back and forth from one direction to the other without losing context.

Stops Explorer

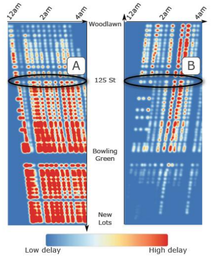


Fig. 7: Stops Explorer: This visualization shows the delay at specific stations for subway line 4, for trips that take place at early hours of the day on weekdays. Regions A and B show shared behavior for uptown and downtown trips, with increase in delay at (or right after) that station, likely due to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

Speed

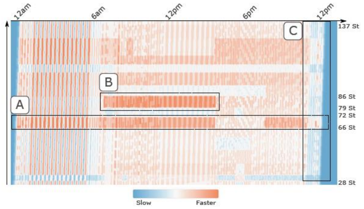


Fig. 8: Speed visualization with Trips Explorer for uptown trips in subway line 1. Region A shows that the speed for trips between 66th St and 72th St stations is mostly constant, except during peak hours,when vehicles run slower. Localized regions depict vehicles running faster than usual (region B) or slower during late nights (region C).

An Overview of TR-EX Interface

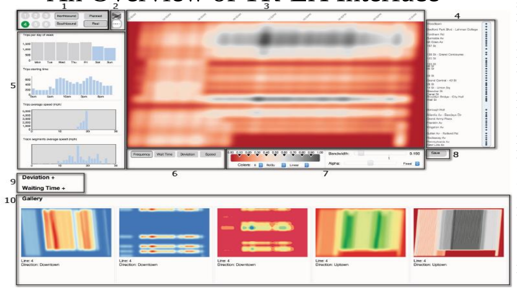


Fig. 9: Overview of the TR-EX proof-of-concept prototype.

Box1:data selection panel
Box2:.The selection between Trips Explorer and Stops Explorer
Box3:main plot
Box4:The Subway Station list
Box5:The Filter Panel
Box6:The attributes selection
Box7:KDE selection
Box8:The save button to save a particular configuration to the Galleryshown in box 10
Box9:additional summary plots for specific attributes selection
Box10:Gallery

Wait Time:

Wait time is computed as half headway, assuming a uniform distribution of passenger arrivals at stations for high-frequency transit services.

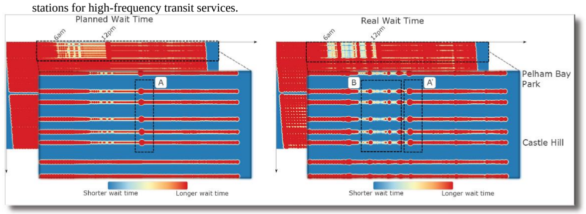


Fig. 10: Comparing planned (left) and observed (right) wait times at stations for subway line 6 toward Brooklyn Bridge - City Hall on weekdays.The visualizations show a substantial divergence between the planned and actual wait times in the highlighted region (top). Zooming into this region, we can see 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.

Reliability

CV = standard deviation/mean actual wait time
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:

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

CV

Reliabillity

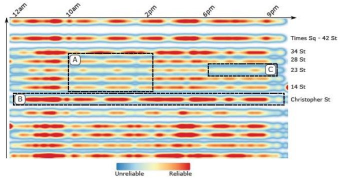


Fig. 11: Reliability visualization of weekdays trips for subway line 1 toward Van Cortlandt Park 242nd Street. Region B confirms the hypothesis that stations with lower demand are more reliable, and region C shows that peak hours cause considerable perturbations in wait time due to higher frequency of vehicles, and resulting in lower reliability. Region A in Chelsea presents unexpected low reliability between 10am and hours.

Comments/Critiques

1.In order to build the real-service dataset, we need to use real-time data provided by the transportation systems,the systems used to capture control signals and train movement can fail.

2.It is good to validate design decisions by interviewing with a transportation expert and doing case studies.However,this is far from enough.Further research needs to be done.

3.It will be better if a map view is intergrated into TR-EX system.

4.A number of features could be added,like the stop that experiences the most delays.

5.For a web-based approach,it is quite difficult to get consistent UI interaction and design for different browsers.

6.As is stated in the interview with a transportation expert,KDE might be confusing to its target users who lack high-level visualization literacy.

7.Computing and displaying the differences directly instead of looking at two different visualizations placed side by side to identify the differences between planned and actual wait times can be much better.

8.TR-EX can be applied to other areas like buses and trains.

Any questions or comments?

Thank you!