Drinking from the Fire Hose: Visualizing Metrorail’s Fare System Data

Justin D. Antos, Corresponding Author
Office of Planning
Washington Metropolitan Area Transit Authority
600 Fifth Street, NW, Washington, DC 20001
Tel: 202-962-2535; Fax: 202-962-1409; email: jantos@wmata.com

Michael Eichler
Office of Planning
Washington Metropolitan Area Transit Authority
600 Fifth Street, NW, Washington, DC 20001
Tel: 202-962-2020; Fax: 202-962-1409; email: meichler@wmata.com

Word count: 2,858 words text + 6 figures x 250 words (each) = 4,358 words

Submission Date: July 30, 2014
ABSTRACT
Every day, rail customers on the Washington Metropolitan Area Transit Authority’s (WMATA, or Metro) Metrorail system generate a vast amount of data on their interactions with the transit network. Metro’s fare system records riders’ travel patterns, journey times, and much more - at a very high level of detail. As a result, Metro is challenged to make sense of this “big data” and display it in ways that are useful to decision-makers.

This paper demonstrates several visualizations - built using data from WMATA’s fare system - that the Authority is using to learn more about its customers, measure performance, inform policy, and understand long-term trends.

Keywords: Data Visualization, Fare System, Farebox, Visualization, Ridership, Transit
INTRODUCTION TO WMATA AND ITS FARE SYSTEM DATA

The Washington Metropolitan Area Transit Authority (WMATA, or Metro), operates Metrorail, Metrobus, and MetroAccess paratransit services in the Washington D.C. region. The Authority provides approximately 1.2 million trips each day throughout seven jurisdictions in the District of Columbia, Maryland, and Virginia.

Metrorail’s fare structure fairly complex, and thus from an analysis perspective, the data is fairly rich. Fares vary by distance traveled and time of day, so customers are charged based on where and when they enter and exit the rail system. The distance-based fare system requires customers to process their fare media at both entry and exit fare gates. When rail passengers “tap out” as they exit, the fare system records a complete trip with origin, destination, and travel time.

Most riders pay fares by tapping a SmarTrip card against an electronic target at the turnstiles. Rail riders may also use a paper magnetic card, and bus riders may also pay in cash as they board, however SmarTrip usage was over 90% of transactions in 2013.

Riders using a SmarTrip card receive a discount when transferring between Metrorail and Metrobus, and bus-to-bus transfers are free. Data generated by use of the fare system can therefore be mined to identify full linked trips with first leg, transfer point, second leg, and so on.

CHALLENGES IN MINING FARE SYSTEM DATA

Metro’s primary challenge in summarizing and analyzing fare system data is that it can be like trying to drink from the proverbial fire hose. The fare system reports vast quantities of information, and different dimensions over which to display the information, that no single visualization or diagram can capture the depth and breadth of the “big data.” For example, rail data summarized by origin and destination across 86 stations quickly yields 7,396 (86 × 86) data points, which is difficult to depict visually. Raw transaction-level data covering all rail riders for 3 months (which is only partially visualized near the end of this paper) involves mining a table of over 48 million data points.

WMATA’s fare system data is assembled into a datamart which reports ridership counts, revenue totals, and travel times (rail only) along a variety of dimensions, including:

- Rider class (student, senior, full fare, employee, etc.)
- Rail origin and destination station, mezzanine
- Rail entry and exit time interval (day, period, hour, half-hour, 15-minute interval)
- Rail service/schedule operated at the time (weekday, modified Saturday, etc.)
- Bus route, line, direction
- Fare medium (SmarTrip, paper farecard, cash)
- Transfers applied (bus-to-rail, rail-to-bus, bus-to-bus)
- Source of fare paid (employer-provided pre-tax benefit, federal benefit, etc.)
- Parking usage transactions by station, exit lane

In addition, WMATA’s fare system also records a wealth of data on “sale” transactions where customers purchase fare media and/or load money onto their SmarTrip cards. However, this paper focuses on leveraging data describing how customers use transit services.
SIX YEARS OF METRORAIL RIDERSHIP, BY DAY

As a primary indicator of a transit agency’s success, ridership is scrutinized and reported in many ways. However, ways of aggregating and comparing ridership numbers (vs. last year, vs. budget, etc.) can make it difficult to see long-term trends or natural variations in the data. In addition, ridership data is so rich that it can be visualized in countless ways.

To look at ridership in a new way, Figure 1 shows total Metrorail ridership for every single day since 2008. Each day is shown as a single cell, and each day is colored according to the ridership on the day - with darkest red being the smallest range (0 to 99,999) and the darkest green being the highest range (greater than 1,000,000). This visualization was designed for viewing in an online environment (1).
This visualization was created using the D3 Javascript library and was inspired by and adapted from Bostock (2). The figure shows 3,000 individual points of data representing over one trillion rail trips over six years illustrated above.

Visualizing Metrorail ridership like this shows several trends:
• Ridership drops on weekends compared to weekdays, and Saturdays are typically higher than Sundays.
• Mondays and Fridays are usually slightly lower ridership than midweek.
• Red-colored weekdays are often holidays such as Memorial Day, and Thanksgiving Day.
• Metrorail ridership shows a distinct seasonal pattern – reaching peak usage in late March/early April for the National Cherry Blossom Festival in downtown Washington D.C., and then again during the summer months when tourism increases.
• Special events cause anomalies in ridership. For example, President Obama’s Inauguration in January 2009 (the darkest green) was Metrorail’s highest-ridership day ever at 1.2 million trips. Severe snowstorms in February 2010 caused many schools and workplaces in the D.C. region to close for several days, dramatically reducing Metrorail ridership – note the string of dark red weekdays at that time.

METRORAIL RIDERSHIP ON HOLIDAYS
For about ten days each year, federal holidays fall on a weekday. Some of these days are holidays for many commuters, but not for all. While nearly all holidays have lower ridership than a typical weekday, understanding the patterns of demand for these days can help Metro decide appropriate service levels, anticipate needs, and plan weekend trackwork.

Figure 2 digs into several individual days and shows systemwide rail ridership by half-hour interval. The lines representing holidays reflect the average of three years of holidays (2010, 2011, and 2012), averaged across the 30-minute interval. For reference, a typical weekday is shown in brown dashed line, and a typical Saturday is shown as black dotted line.

![Figure 2: Metrorail Ridership by Time of Day on Holidays](image)

**FIGURE 2** Metrorail Ridership by Time of Day on Holidays

This visualization (Figure 2) shows that not all holidays are equal. Columbus Day and Veterans
Day are somewhere in between a weekday and a weekend. These two holidays show moderate peaks at typical rush hours, almost like a miniature weekday. This could be because most federal workers (who account for over a third of peak Metrorail ridership) get these days as vacation, while private-sector employees are more likely to commute to work. By contrast, Memorial Day behaves more like a Saturday or Sunday, while ridership on Thanksgiving Day is very low.

In response to these unique demand patterns, Metro adjusts rail schedules to meet demand. On holidays like Memorial Day and others, Metro operates on a modified Saturday schedule, with a 5:00am opening and a midnight closing. On Columbus Day and Veterans Day however, Metro operates on a modified “Saturday Holiday” schedule, providing extra service beyond Saturday levels during peak hours. All holiday rail schedules are modified to accommodate scheduled track work, if necessary.

ORIGIN OF PARKING CUSTOMERS

WMATA operates parking facilities at 35 of its 86 Metrorail stations, and parking customers account for approximately 15% of total ridership on typical weekday. Located primarily at stations outside the downtown core, these parking lots and garages are typically adjacent to, or a short walk from, the station entrance. Customers must tap their SmarTrip fare card or credit card to access the parking facilities and the fee is approximately $5.00/day. Metro provides this parking to give customers from across the region a way to access Metrorail, and at over 60,000 spaces Metro is one of the largest parking providers in the region. Some stations have relatively small lots of 200-300 spaces, while larger end-of-line stations host over thousands of spaces in multiple garages.

Understanding the origin of Metro’s parking customers can give insight into what kind of market is being served, the geographic extent of Metrorail users, and even the strategic role of parking and transit-oriented development.
To construct Figure 3, all SmarTrip parking transactions for a typical weekday in 2012 were tallied, and data was filtered to show only SmarTrip cards with a registered address, and address of the card was geocoded. To make the map more legible, and to accurately represent concentrations of customers in dense residential areas, a half-mile (0.8-km) square grid (using the Create Fishnet function in ArcGIS) was overlaid on the region, and customers were aggregated into grid cells. The resulting visualization shows about 5,700 grid cells representing over 45,000 parking transactions.
Several patterns emerge from Figure 3:

- Some parking customers travel very long distances before parking at a Metrorail station – as far away as Baltimore, Dale City, and Frederick.
- A sizeable share of parking customers (13%) live outside the WMATA Compact jurisdictions, so do not contribute to the local subsidies Metro receives.
- There is no “donut” of white space around the stations themselves, apparently indicating parking customers who appear to live very close to the rail station.

The phenomenon of short-distance parking customers was investigated further. These could signal poor pedestrian and bicycle connections to the station area, and may represent an opportunity for pedestrian and bicycle access improvements. To quantify the actual distance between customers’ home origin and the station, the straightline distance was estimated using the Near geoprocessing function in ArcGIS.

The results showed that systemwide, one-third of Metro parking customers drive from less than three miles (4.8 km) away to their station. Eight percent of customers drive from less than one mile (1.6 km). Short-distance parking customers are particularly pronounced at some stations, where the parking facilities appear to serve Metrorail customers arriving from little more than the immediate surrounding neighborhoods. At Forest Glen station, for example, 67% of parking customers live within two miles (3.2km) of the station. Van Dorn Street, West Hyattsville, and Fort Totten, for example, are especially local, with more than 30% of parking customers hailing from less than one mile from the station entrance.

ORIGIN OF PARKING CUSTOMERS, BY STATION

Missing from Figure 3 is information on which stations customers are parking their car. It is difficult to discern the market area or “commute shed” of each station’s parking facilities.

To address this, parking customer origins are again visualized, but the dominant station (the station attracting the highest number of customers) is shown for each grid cell as a distinct color (Figure 3). The color of a station’s parking shed matches the color of the dot representing the station. While this technique might oversimplify a square containing, say, 29 customers driving to Station X and 30 customers driving to Station Y, it does give a broad sense of each station’s market area.
FIGURE 4 Origin of Metrorail Parking Customers, by Primary Station

Stations’ market areas tend to follow commuter highways and arterial roadways leading to the downtown core. The commute shed for West Falls Church in yellow, for example, stretches west along the Dulles Toll Road.

This visualization also demonstrates that the end-of-line stations dominate the long-distance drive-to-metro markets, while the rest of Metro’s parking facilities serve small market areas.

If one purpose of commuter park-and-ride facilities is to extend the geographic reach of transit, only a handful of Metrorail stations are actually achieving this. End-of-line stations with large
parking facilities, representing only 25% of Metrorail stations with parking, cover 70% of the total parking commute shed in Figure 4.

The other 26 stations with parking are primarily providing neighborhood parking, and could represent opportunities for transit-oriented development. Addressing the need for short-distance auto access to Metrorail by other means could allow WMATA to grow ridership beyond the capacity constraints of the parking lots. If Metro could shift some of these short auto trips to another mode, it may be able to convert land occupied by parking lots into higher and more valuable uses.

**MAPPING THE IMPACT OF DISRUPTIONS ON RAIL CUSTOMERS’ TRAVEL TIMES**

Like any rail system, road, or transportation network, Metrorail operations are sometimes disrupted. Equipment fails, emergencies arise, and passengers are delayed. Even after an incident is cleared up and normal operations resume, residual delays can continue for customers. Metro actively tracks and reports the number of trains that do not adhere to headways – i.e., that arrive outside of a given window of time, and the number and duration of train delays (3). But while Metro has a good understanding of its train movements, it is more difficult to understand the impact to passengers – how many passengers were impacted, and by how much.

Fare system data can help fulfill this need. When Metrorail customers exit the rail network, the fare system records the time elapsed between “tap in” and “tap out.” By mining this data, Metro can better understand how travel times change during a disruption.

![Travel Time, Shady Grove to Farragut North](image)

**FIGURE 5** Average Customer Travel Time for One Station Pair, by Day

Figure 5 demonstrates that travel time data from the fare system appears to be suitable for such an analysis. The average travel time for morning commuters between Shady Grove and Farragut North stations is typically around 40-42 minutes, but on several mornings that travel time increases by 10 minutes or more due to disruptions.
However, the best way to visually display this data remains a challenge. Figure 5 shows travel time averaged across all travelers between two stations for an entire AM Peak period, which masks a great deal of variation across the roughly 1,400 passengers who make this particular journey, and across the four-hour timeframe. In addition, each Figure 5 shows only the impact to riders on one station pair, which represents less than 1% of the total ridership in the AM Peak.

Figure 6 attempts to address some of these challenges, by extending the visualization to all origin-destination pairs. (This figure shows several frames of an animation, which in part helps reveal variation of impacts across time intervals). In Figure 6, every origin-destination pair is represented as a single line overlaid on the schematic Metrorail system map, and the width of the line denotes ridership volume. The color of the line represents how much the actual travel time on one day differed from the average travel time across a three-month period – darker red indicates above-average travel times. For legibility, station pairs with ridership below 100 customers per day are not shown.

In the end, each panel (day) of Figure 6 summarizes the travel times of over 220,000 customers in the AM Peak, visualized using approximately 2,000 lines. This visualization was constructed using Tableau Desktop, leveraging the “Path” shelf to visualize origin-destination activity (4), and based on a direct Oracle connection to Metrorail fare system data.
FIGURE 6 Map of Actual vs. Average Travel Time, by station pair, four sample morning commutes.
Visualizing Metrorail travel times in this way shows several findings. First, it partly confirms that the fare system data is recording customer impacts. For instance, all stations along particular lines appear to change color as a group, reflecting that a disruption on one portion of a line usually reverberates down the line. The “cascading” effects of delay are also evident, where the effect of disruption on one line follows customers who transfer between lines.

Second, Metrorail fare system data can form the groundwork for quantifying passenger delay. With an understanding of how many customers are impacted by an incident, and by how much, Metro can see how operational responses can impact customers, and how to best communicate with customers about expected travel times.

CONCLUSION
Metro is learning a great deal about customer travel patterns, travel times, markets, and more by mining and visualizing data from its fare system. Since each new day brings Metro over a million data points organized across hundreds of dimensions, the primary challenges in making sense of this data is its sheer size and complexity, and crafting the best questions to the data source. This paper has shown how analytical techniques from “big data” can help distill the data into formats meaningful for decision-makers and non-analysts. The visualizations presented in this paper are helping the Authority to understand more about its customers, measure performance, inform policy, and understand long-term trends.

DISCLAIMER
The contents of this research reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington Area Metropolitan Area Transit Authority. This report does not constitute a standard, specification or regulation.

REFERENCES