Understanding Road Usage Patterns of Hubway Riders

INTRODUCTION

While much work has been done in the transportation community on understanding auto trips and road usage\(^1\), alternative modes of transportation such as cycling have received comparatively little attention. Not surprisingly, data on cycling trips is sparse and collected through surveys, which are time intensive, expensive, and limited in sample size. The release of station-to-station trip data in 2012 for Hubway’s first 14 months of operation, however, provides an opportunity to analyze Hubway ridership patterns, and perhaps more generally, cyclist behavior in and around Boston.

By segmenting Hubway trips by the home location of its riders, distinct spatial and temporal patterns in the usage of the system emerge. This analysis explores the similarities and differences across 7 communities that are key contributors to Hubway ridership—Central Boston, Fenway/Kenmore, South End, Back Bay, Allston/Brighton, Cambridge, and Brookline. Beyond station-to-station trends, road usage patterns are inferred by linking road segments to their contributing station pairs, and by extension, the demographic characteristics of riders. This connection between road usage and user attributes provides insight into where and how various subgroups of riders interact with the city. From a business perspective, Hubway could benefit by better understanding its existing and potential markets. Furthermore, awareness of roads with heavy bike traffic may be helpful to public agencies planning for new or improved bike facilities.

DATA

Hubway launched its bike-sharing operation in July 2011 with 600 bikes at 61 stations across Boston. In July 2013, the system had grown to 1,100 bikes at more than 110 stations and celebrated its one millionth trip\(^2\). As part of a data challenge\(^3\), Hubway released station-to-station trip data for its first 14 months of operation (7.28.2011-10.1.2012). This data includes trip characteristics—start and end station, date and time—as well as user attributes, if provided by a registered user, including their gender, age, and home zip code. Station level data includes the name and coordinates for the 95 stations active during the 14 months.

In addition to the Hubway data, Geographic Information System (GIS) data of road and bicycle facilities is used to estimate a trajectory between each station pair. The Massachusetts Department of Transportation maintains Road Inventory\(^4\) and Bicycle Facility Inventory\(^5\) GIS layers containing the spatial linework and attributes for roadways and existing bike facilities in Massachusetts, respectively.

METHODS

Due to the incomplete nature of the Bicycle Facility file, the network used in this analysis is a subset of the Road Inventory network, while the Bicycle Facility network is merely used as a reference. The road network was developed by selecting arterials in the study area that could be potential bicycle paths. After this preliminary network was selected, all Hubway trips were assigned to the network using a shortest path approach. Links that were not included in any of the shortest paths between station pairs were then eliminated, producing the final network shown in Figure 1.

Figure 1. Hubway road network (grey) with Hubway stations (red)

Because cyclists’ may prefer a scenic or low-traffic path to the shortest one, using shortest path assignment could be called into question. Furthermore, due to the nature of bike-sharing, some riders may check out a bike for a leisure ride rather than a direct trip. Accordingly, this methodology is validated using median trip duration from the Hubway data.
For each station pair with a sufficient number of trips (20), the median trip duration $a_{ij}$ was calculated from all trips made by ‘Registered’ subscribers. By using the median time of trips made by the subscriber class that is most likely to make explicit trips between origins and destinations, this duration best captures the time it takes to make a direct trip between each station pair. The time for the assigned shortest path between each station pair $b_{ij}$ was then calculated using the distance and an assumed cycling speed of 9 mph. For every station pair, a ratio of the trip duration from the data and assignment $D_{ij} = a_{ij}/b_{ij}$ is calculated and plotted in Figure 2 relative to (a) trip distance and (b) station pair volume.

Figure 2 demonstrates that the majority of station pairs cluster around $D_{ij} = 1$, with the worst outliers occurring for short distance trips and station pairs with few trips, as expected. Overall, this method reasonably approximates the time between each station pair and by extension the assigned path. With this procedure, a network of road usage is developed, with road segments represented by 1,140 edges and intersections by 714 nodes. The number of trips using each road segment from node $i$ to node $j$ is analogous to an edge weight, $F_{ij}$.

With the development of the road network, spatial patterns emerged, revealing sub-networks with distinct ridership patterns and user characteristics. These sub-networks are defined by aggregating zip codes into communities that represent key neighborhoods or towns contributing to the Hubway system, as shown in Figure 3.

**RESULTS**

With the Hubway trips segmented by community, ridership trends—both in trip characteristics and rider demographics—are clear. One such finding is the difference in the distribution of trips among station pairs across each community. Figure 4a shows for each station pair, the number of trips indexed by the number of trips between the top station pair. This reveals that compared to the other communities, the top station pairs dominate ridership for Central Boston and Allston/Brighton riders.

By contrast, Brookline and Cambridge ridership is spread more evenly across station pairs. In terms of road segments, Figure 4b illustrate a steep drop off in segment volume after the top 100 segments or so. The fact that a small number of roads are used for the majority of trips indicates that most of the trips made by riders of a given community use a small portion of the road network.

---

Figure 2. (a) LEFT: The ratio of trip duration from the data and assignment $D_{ij}$ vs. trip distance for each station pair. (b) RIGHT: The ratio of trip duration from the data and assignment $D_{ij}$ vs. trips for each station pair. The red line ($D_{ij}=1$) reflects consistent median trip duration and the duration of the shortest path.

Figure 3. Key ridership communities and their zip codes. Hubway stations are shown in black. Total Hubway trips are presented for each community.
The Hubway trip distributions in Figure 4 indicate that most trips use relatively few road segments, but how does this look spatially? Figure 5 illustrates the spatial networks of each of the 7 communities. As expected, the road segments and stations with the heaviest bike traffic for a given community are located near that community. Beyond just a visible center for each sub-network, the breadth of the trips is notably different across the communities. While Central Boston, the South End, and Back Bay are relatively centralized, main corridors are evident for Allston/Brighton, Brookline, and Cambridge. Another key pattern is the lack of trips in and out of Cambridge from users in the other communities, whereas Cambridge residents seem to travel to the other communities much more.

These spatial trends suggest there may be notable differences in the trip duration and distance across residents of the different communities. The distribution of trip durations is consistent with this theory, as shown in Figure 6. As expected, the duration of trips is closely related to the distance of communities relative to Central Boston. Trips by riders living in Central Boston are shorter than for the other communities, followed by the other nearby communities of Fenway/Kenmore, the South End, and Back Bay. Cambridge, Allston/Brighton, and Brookline, on the other hand, are generally longer. This suggests that, especially for Allston/Brighton and Brookline, riders are generally not using the system within their own community but rather traveling into
Boston. This pattern is consistent with the corridors visible in the spatial networks of Figure 5.

These observed variations beg the question—are trip distribution and duration different because of the geography of the system and area or do other differences contribute? As Figure 7a shows, demographics may play an important role in rider behavior. The distribution of rider age reveals that Allston/Brighton, Fenway/Kenmore, and Cambridge have considerably younger riders, as you may expect for areas with large student populations. The South End and Back Bay, on the other hand, are home to expensive real estate, and accordingly, trend older with professionals and retirees. In Figure 7b, rider age is linked to road segments. Deeper saturation shows areas with older riders, namely Central Boston, Back Bay, and the South End. Consistent with the overall distribution in Figure 7a, roads in Cambridge, Allston/Brighton, and Fenway/Kenmore are used by younger riders.

One would expect that behavioral differences go hand-in-hand with lifestyle differences of different demographic groups. As such, temporal patterns are compared across the communities in Figure 8. Patterns again differ for Allston/Brighton, Fenway/Kenmore, and Cambridge, which exhibit a larger midday peak on weekdays than the other communities, suggesting riders are less prone to follow a typical commuting pattern, consistent with what you would expect for areas with sizeable student populations. Similarly, riders from these communities make more trips later on weekends and in the early morning hours compared with Back Bay, South End, and Central Boston users.
CONCLUSIONS & OPEN QUESTIONS

When segmented by user's home location, Hubway ridership exhibits definite spatial and temporal patterns. With demographic characteristics of the users, such as age, we are able to gain insight into reasons for these differences in addition to those explained by geography. Beyond statistics relevant to Hubway station pair ridership, this methodology connects user attributes with individual road segments, allowing for a deeper understanding of where and how cyclists use a city.

To expand on this work in the future, data on additional demographic characteristics such as per capita income or car ownership could be incorporated to explore further reasons behind the cyclists' behavior. Additionally, a closer comparison of the high-traffic bicycle corridors with bicycle facilities would provide important information to mitigate cyclist safety concerns. Lastly, an interesting exercise would be to compare Hubway ridership patterns with patterns of other Boston area transportation modes, namely auto, bus, and subway.

---


2 http://thehubway.com/news/2013/07/30/hubway-celebrates-two-years-and-one-million-rides-

3 http://hubwaydatachallenge.org/

4 http://www.massdot.state.ma.us/planning/Main/MapsDataReports/Data/GISData/RoadInventory.aspx

5 http://www.massdot.state.ma.us/planning/Main/MapsDataReports/Data/GISData/BicycleInventory.aspx