The Role of Bluebikes in Boston Commute Patterns

Meia Alsup, Charles Chimento, Jinghong Lyu, Miranda McClellan, Yue Wu
6.439 - Massachusetts Institute of Technology
Professors Stefanie Jegelka and Caroline Uhler
(malsup, cchiment, jlyu, mcclm, yuewu) @mit.edu

Abstract— Roughly 500,000 people use the Boston public transit systems every week-day to commute. Since coming to Boston 2011, Boston's Bluebikes system has offered residents a new option for their daily commute. In this paper we analyze how the Bluebikes System is used for commuting. Specifically we investigate seasonality, time-saved versus other commute methods, and importance and popularity of different routes and nodes in the network. We find that the Bluebikes system supports thousands of commuter trips per day at its peak during the summer, but this falls off to a negligible amount during the winter months. Further, we find that the Bluebikes time efficiency advantage peaks for trips about 10-20 minutes long.

I. Introduction

Boston was among the first large cities in the U.S to commit to a bike sharing system. First established in 2011, the number of bike stations has grown from 60 to more than 190, and the bike quantity from 600 to 1,800. The citys public-owned bike share system Bluebikes (originally Hubway, and now owned by ride-sharing conglomerate, Lyft) has changed the way citizens commute. Since launch, riders have taken over six million trips with more than 1.3 million trips in 2017 alone. Through an agreement with the city of Boston, the ridership data is publicly available, providing insight into the commute patterns of Boston's population.

In 2012, the Bluebikes system expanded into the neighboring municipalities of Brookline, Cambridge, and Somerville. Ridership has increased as the Bluebikes system has expanded. Local universities such as Harvard and MIT actively support the adoption of these shared bikes by offering discounted yearly subscriptions. In this paper, we study the common routes taken between the 195 existing Bluebikes stations in the greater Boston urban area to understand how the bike share system has changed and shaped Bostonians' commuting trends. We also consider how these observed trends might inform public transit policies such as tailoring MBTA transit services to the drop in Bluebikes' usage during the cold winter months. Ultimately, this investigation offered insight into citizens' commute modes across different jurisdictions in Boston and evaluate their efficiency in terms of time.

This paper contributes 3 main elements to our understanding how citizens use the Bluebikes system for transit during peak commuting times (7am-11am and 3pm-7pm):
1) Understanding monthly, daily, and hourly seasonality trends in aggregate as well as the variation across stations through time series analysis,
2) Estimating time-saved from biking relative to walking, using public transportation or private car,
3) Identifying similar station and trends in most popular routes taken on Bluebikes using network analysis, and clustering.

From these analyses, we also make recommendations for additional routes or resource allocations for public transit that would benefit commuters.

II. Related Works

The rapid emergence of bike sharing systems in large cities in both Europe and the US has been accompanied with a rapid rise in papers and research studying the prevailing trends in these systems. A sizable branch of this research seeks to understand who is adopting bike-sharing, the motivations for doing so, and the causal effects on other modes of transportation.

Kabra et al. uses the dominant bike-sharing system in Paris to estimate the impact of two facets of system performance on bike-share ridership: accessibility to bike stations (the distance that users must walk to reach the stations) and bike-availability (the likelihood of finding a bicycle) [2]. The study finds that every additional meter of walking to a bike station would reduce a user's likelihood of using a bike from that station by $0.194\%(\pm 0.0693\%)$, and a 10% increase in bike-availability increases ridership by $12.211\%(\pm 1.097\%)$. This study implies that planners should seriously consider bike station locations to promote ridership, and reveals time-efficiency as a key motivation for users of bike-sharing.

What better place to introduce more time-efficient practices than work or school commute times? In ”Bikeshare: A Review of Recent Literature”, Elliot Fishman finds that in Washington, D.C. 43% of long-term members report that their last trip was work related, compared to 2% for short-term users. Similarly, in London, 52% of annual subscribers report that their last trip was commuting to or from work [3]. This paper concludes that commuting is the most common trip purpose for annual members. This conclusion will be tested in this paper to see if Boston's Bluebikes users are predominately using the bikes for work related commute.

In a parallel vein, Kyle Gebhart and Robert Noland demonstrate that the Metro is used as a back-up when the weather is unfavorable for biking [4]. The results provide data to support the intuition that users who have adopted bike-sharing are liable to reverting to public transportation in poor weather conditions.

Another study led by Elliot Martin and Susan Shaheen identifies the types of people that are switching away and towards public transportation in response to bike-sharing [5]. Results suggest that the introduction of bike sharing causes inner-city residents to shift away from public transportation, presumably in favor of the newly available bikes.

III. Data

For our analysis, we use the publicly available open data set provided by Bluebikes. The open data set includes all trips taken by riders since the beginning of the program [1]. The data set provides various metadata associated with each ride as shown in Table III.

In this investigation, we limit our data of interest to rides that occur during a peak transit time in the morning (7am-11am) or in the evening (3pm-7pm). We also only consider rides taken by subscribers. We ignore single trip riders because they are more likely to be tourists or one-off users rather than commuters. This is because the subscription payment model favors high frequency users such as commuters [3]. While not all of the riders who subscribe are guaranteed to be commuters, trends in commuting will be easier to identify in the subscriber data.

We consider data from October 2017 until September 2018, the most recent calendar year of data available. There are a total of 1,688,523 rides in the data set of which 1,626,368 have different start and end destinations. Namely, the other rides and start and end at the same place and are not relevant to an analysis of commuting trends. Of the rides with different start and end destinations, 1,353,960 of the rides were taken by those with a subscription to the Bluebikes service. We also note that there is no user id associated with the public data set, which means back attributing commuting rides cannot be done via patterns on a per-user basis. The Bluebikes data set is already cleaned of false starts in docking or un-docking bikes and employee use for maintenance.

We also used data collected from the Google Maps Distance Matrix API to estimate travel time between places in the city. From the API, we obtain the travel time between the specified start and endpoints.

Category	Data Type
Trip Duration (sec)	int
Start Time	Date Time
Stop Time	Date Time
Start Station	ID and Name
Start Station Location	Longitude and Latitude
End Station	ID and Name
End Station Location	Longitude and Latitude
Bikeid	ID
User Type	”Subscriber” or ”Casual”
Birth Year	Date
Gender	0/1/2 for NA/M/F

IV. Preliminary Analysis

An initial inspection of the data suggests that subscribed users rely on Bluebikes for commuting to work. First, we analyze how the subscribers might be different in how they use the bikes than casual pay-per-use customers (Figure 1). The casual user seems to be more representative of a tourist or ”joy-rider”, whose usage rates are high in the summer months and converge to almost zero in the winter months with no long term trend. Further distinguishing is that their hourly usage rates are relatively constant throughout the day, whereas the subscriber usage rates vary with peak work-time commute patterns. Since such a user is less concerned with time efficiency of travel, they are excluded from the data analysis when the interest is comparing route efficiency for typical commute routes.

The visualization of the trend since 2015 exhibits a low daily mean and variance in the winter, whereas the summer months reveal a much stronger influence of weekend versus weekday on ridership. Removing

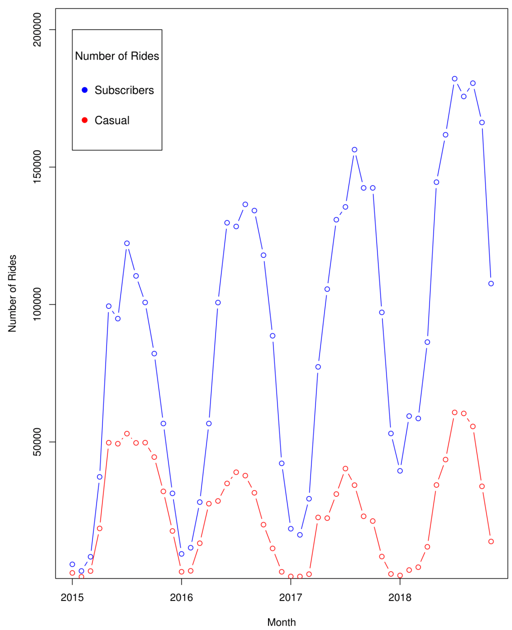


Fig. 1: Subscribers' usage shows a growing trend, whereas the casual users' only consists of a seasonal component.

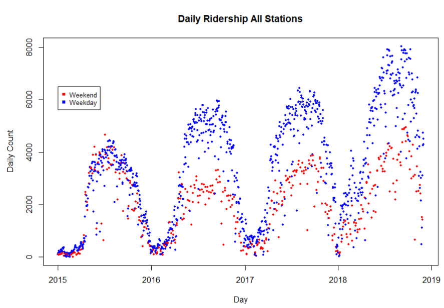
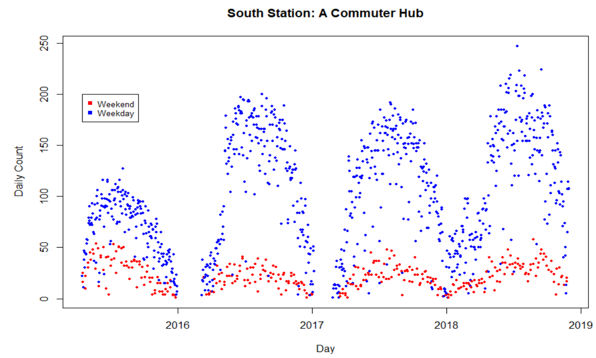
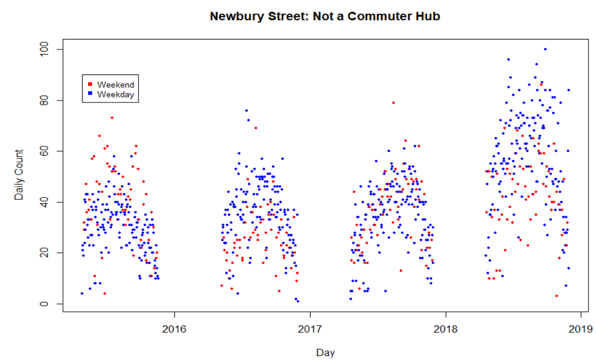


Fig. 2: The smaller variance of ridership in the winter suggests that those who ride Bluebikes in the winter are much more consistent with how often they use them.



(a) The missing data during the winter months of 2016 and 2017 is likely due to construction. Note the pronounced gap between weekend and weekday for South Station compared to the data averaged over all the stations in Figure 2.



(b) The usage of the station near the shopping center shows, as expected, little dependency on whether it is a weekday or weekend and is thus representative of the fact that not all nodes are commuter hubs. Note that the mean usage rate is also lower for stations that are not commuter hubs.

Fig. 3: Comparing a station well suited for work-commute and a station in a shopping district

the casual users reveals a more defined relationship between the weekday and weekend ridership as shown in Figure 2.

As shown in Fig 2, daily ridership is higher for subscribed users during the week than on the weekends. This trend is pronounced for popular stations that are well suited for work commute. South Station, one of the most popular stations, demonstrates this trend in Figure 5a. The bike station at Newbury St, a prominent shopping center, is juxtaposed with the South Station station. Whereas South Station is a known work-commute hub and shows strong correlation between weekday and weekend, Newbury St ridership is uncorrelated with weekday versus weekend.

The hourly peak usage times, however, vary among the popular stations in a manner that reflects its use as a hub for work-commute. Peak work-commuter time accounts for over 90 percent of the total daily usage for both South Station and MIT Stata Center. Figure 5 shows the hourly trips originating from the MIT Stata Center and South Station for a typical week in November.

From this, we hypothesize that ridership during our selected peak times is predominately used for weekday work-commute.

Next, we identified the most popular stations for where rides start and end. We plot the number of rides that start and end at each station versus the station latitude and longitude in the heat map in Figure 4. You can make out the Charles River as the small gap in the middle. Key stations are colored red or yellow, and among them are the Amherst Alley at Massachusetts Avenue station on MIT's dorm row, and the docking station outside of the MIT's Stata Center. Is MIT destined to be the center of the bike share commuting world?

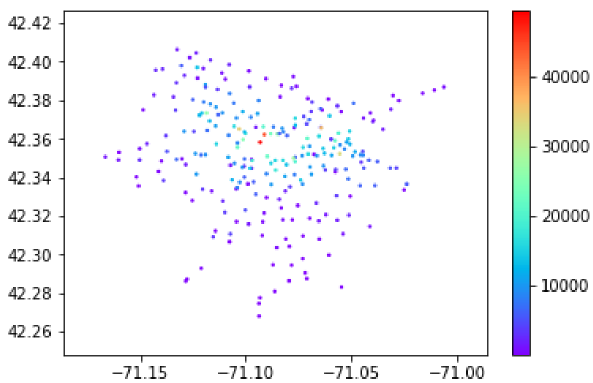
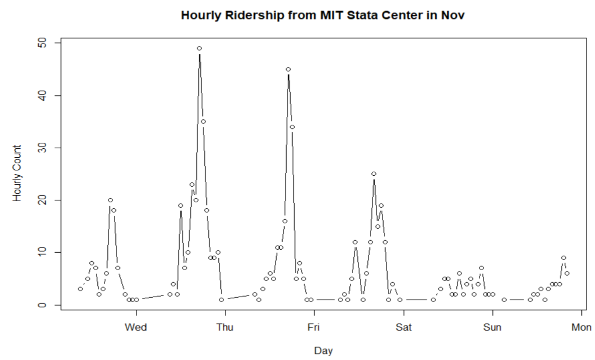
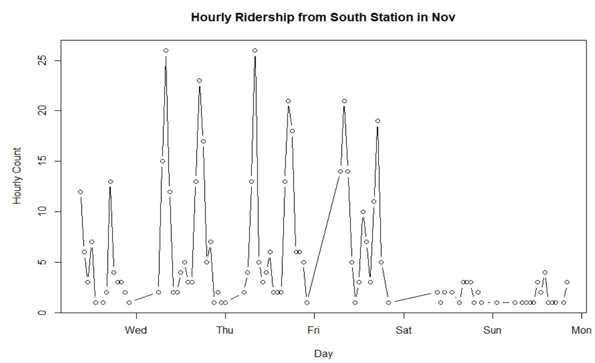


Fig. 4: A heat map of the Bluebikes docks within the city where high intensity corresponds to more trips ending at that dock. The data used is all rides from September 2017-October 2018



(a) MIT Stata Center peaks from 4pm-6pm on Weekdays in June 2018



(b) South Station peaks during both morning and evening work-commute times.

Fig. 5: Whereas the MIT Stata Center show peaks for 4pm-6pm, South Station shows peaks twice daily for 8am-10am as well as 4pm-6pm. Note that rides are attributed to the station of origin.

V. Methodology

We performed two types of analysis on the Bluebikes bike share data set: 1) analysis of time saved by bike sharing to understand the growing popularity and 2) a network analysis to understand popular routes and which stations are similar to each other in terms of use.

A. *Do Bluebikes Save Time?*

To understand time saved, we compare various modes of transportation. We are most interested in the bike routes that are used for commuting, and we proxy for this by analyzing the most popular routes for subscriber type riders. We pull the top five most frequently ridden routes for three time ranges: time spent biking under 10 minutes, in the range of 10-20 minutes, and more than 20 minutes.

As a result, we analyze a total collection of 15 different routes where we see we have 16 different unique start and end points within this set of 15 routes.

In order to understand time saved, we estimate where trips would begin and end on average for each of these routes. We use a Voronoi Diagram to distinguish neighborhoods, where all points within a neighborhood are closest to the bike station that defines the neighborhood.

For each of the 15 routes, we sample 200 pairs of start and end latitude, longitude coordinate pairs. These are sampled from a two-dimensional geometric Gaussian Distribution with the underlying assumption that users are more likely to use Bluebikes if their travel time to the bike station is smaller: for instance, if we are going to sample the start points around the start station: first, we use the coordinate pairs of start station (x_0, y_0) as the means of the samples (x_i, y_i):

$$\mu_x = x_0, \sigma_x = \frac{max\|x_i - x_0\|}{2}$$

$$\mu_y = y_0, \sigma_y = \frac{max\|y_i - y_0\|}{2}$$

Second, the dimensions are scaled according to the

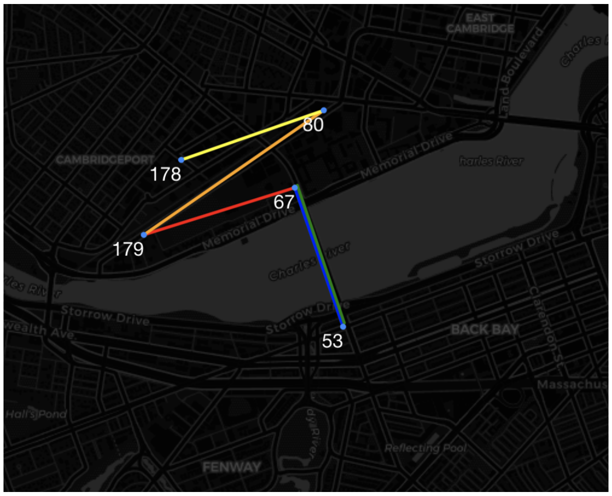


Fig. 6: Five Most popular routes where average bike time is under 10 minutes

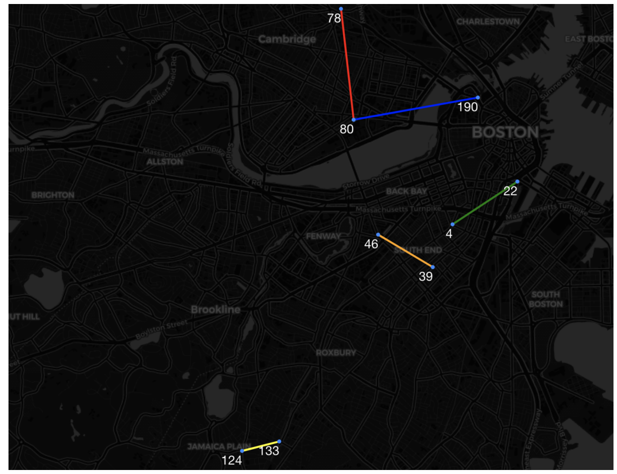


Fig. 8: Five most popular routes where average bike time is over 20 minutes

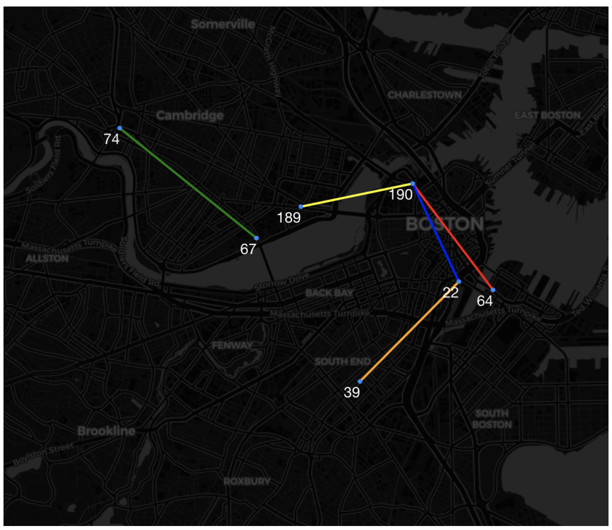


Fig. 7: Five most popular routes where average bike time is 10 - 20 minutes

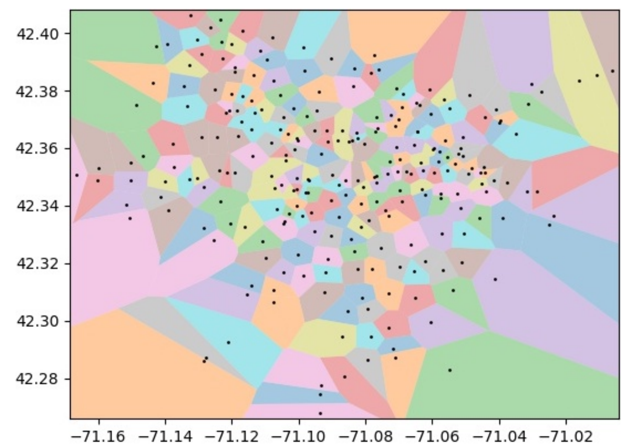


Fig. 9: Voronoi diagram mapping splitting Boston into sections based on the closest Bluebikes station

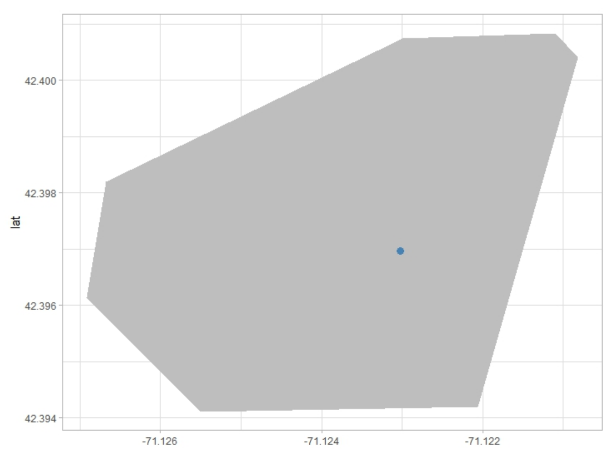


Fig. 10: Using the latitude, longitude coordinate pairs as the center of sampling

σ_x and σ_y so that the furthest corner of a neighborhood has a 5% probability of being selected (be covered within the width of two standard deviations). Third, we use the neighborhoods defined by the Voronoi diagram to cut off the sampled points that are out of the boundaries.

For each of these 200 routes, we pull the walk, transit, and drive time as the baseline of existing transit options prior to Bluebikes using Google Maps Distance Matrix API. This allows us to specify the start and end travel pairs and get the travel time between. To compute the Bluebikes transit time we add the following three components: walk time to the start station, bike ride time, and walk time from the station to the destination.

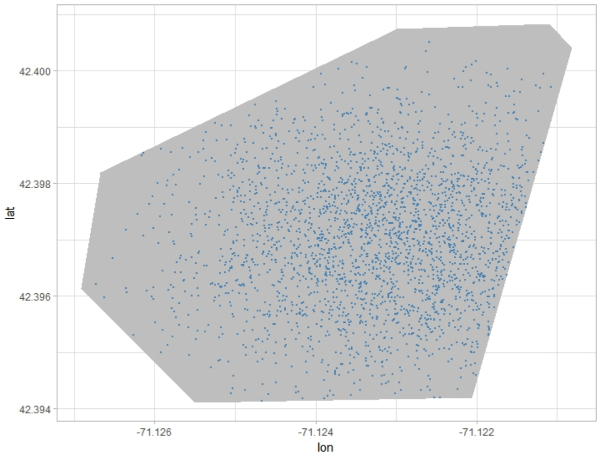


Fig. 11: Sampled start points within the neighborhood

The walk times are also sampled from the Google Maps Distance Matrix API per the Gaussian distribution previously described. Then bike ride time is calculated as the average time from the start station to end station per the Bluebikes data.

B. Which Routes Are Popular?

First, popular routes are identified in the Bluebikes network. Specifically, the popular routes during morning and evening commute times are of interest (07:00-11:00AM and 3:00-7:00PM, respectively). Further, the routes are examined for seasonality. We first split out the trip data during the morning and evening rush hour periods. For both sets, we build a network where edge weights are the number of trips between the starting and ending point of the route. We then visualize the common routes for both periods.

We rely on eigenvector centrality as a proxy for importance in the network. This is done for each month during our study period. For each month, we track the 5 nodes with the highest centralities, and track the seasonal variance and importance of the relationship between the nodes.

C. Which Stations Are Similar?

We wanted to understand if stations across the city displayed similar usage patterns. We represented each Bluebikes dock as an 8-dimensional vector according to the following attributes: location in longitude and latitude, number of trips originating at station, number of trips ending at station, average trip duration in minutes for both ending and starting trips, number of bikes at the station, and average age of people using that station.

We performed Principle Component Analysis (PCA) with three components for the morning subset of data and evening subsets of data individually. The variance from the PCA is explained in Table 3. The results of the PCA vary drastically between the two data sets, which suggest that the station usage or demographics of users differs a lot between the two time periods. Then, using the lower dimension output of the PCA, we performed Spectral Clustering on the two sets of data with 3 clusters. The results of the clustering are seen in Figures 9 and 10 in the 3-space of the principal components. We chose three component PCA because the method explained sufficient variance for each data set (mroe than 80% for morning and 96% for evening commute windows).

Variance in Bluebikes dock usage in the morning commute window is thus described by the three principal components: the duration of trips starting from the dock (42.76%), the duration of trips ending at the dock (23.12%), and finally the number of bikes (14.24%). For the evening, the significant attributes are the duration of trips ending at the dock (87.16%), the number of trips starting at the dock (6.67%) as well as the duration of trips beginning at the dock (3.03%).

We then performed Spectral Clustering on the dimension reduced output of performing PCA. The results of the clustering can be seen in Figures in the results section, we will show the distribution of the clusters on a map of Boston and analyze meaning of intra-cluster similarity between stations.

PC	Morning	Evening
1	.4276	.8716
2	.2313	.0667
3	.1424	.0303
Total	.8013	.9687

VI. Analysis of Results

A. Estimating Time Differences

We find that for the short bike rides (under 10 minutes), only the comparison to walk time makes sense. For the medium and long rides (over 10 minutes) we compare to both public transit and driving time in the analysis.

We also caveat that here we ignore the time that it takes a user to check out and return / lock a bike to the station. We also ignore the time it takes to call / wait for an Uber/Lyft or park a personal vehicle. We also look at only the expected travel time of public

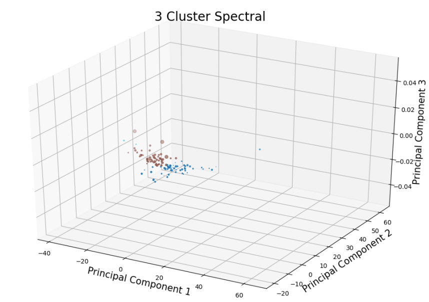


Fig. 12: Results of 3 cluster spectral clustering on the PCA outcome for morning rides from Bluebikes stations in 3 groups represented as red (1), light blue/green (2), and dark blue (3).

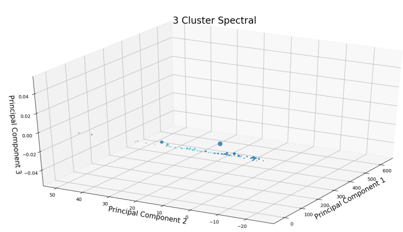


Fig. 13: Results of 3 cluster spectral clustering on the PCA outcome for evening rides from Bluebikes stations in 3 groups represented as red (1), light blue/green (2), and dark blue (3).

transit options, which do not account for the ETA of how far away the next train or bus is. As a result, we suspect all of our values are slight under-estimates, but nevertheless offer unique insights. We do not correct for these since they are very high variance and would add a lot of noise to our results.

For the short trips, we see that walking is often faster than biking. We also note that our Voronoi method of sampling start points might be a little biased in this case. The random sampling method includes routes where both bike stations could be out of the way, in which case the time-saved proposition is less compelling. Short trips in practice likely have higher time saved components since riders only use the bike sharing if the stations are on their route.

For medium length trips, on average our results indicate that biking is faster than public transit, but slower than driving. For the longer trips, biking and public transit are similar and driving is faster.

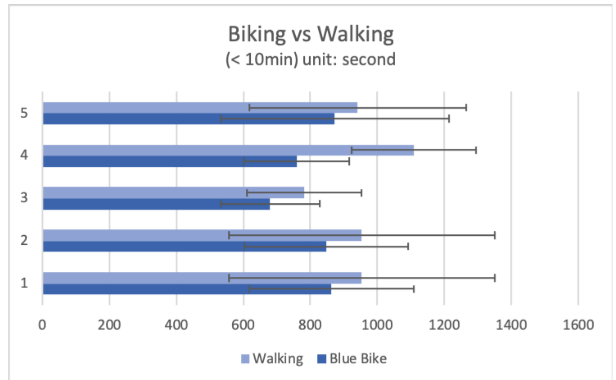


Fig. 14: Transit mode efficacy comparison when bike time is under 10 minutes

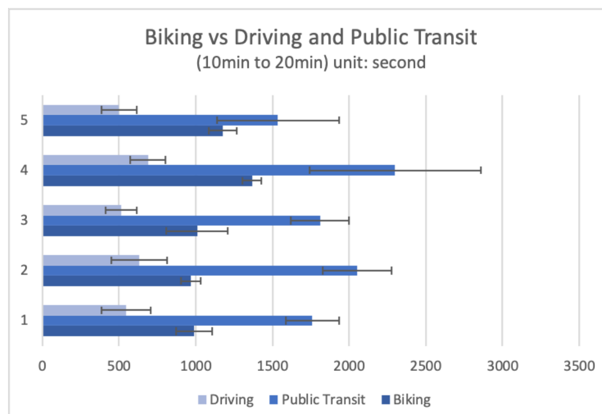


Fig. 15: Transit mode efficacy comparison when bike time is 10 - 20 minutes

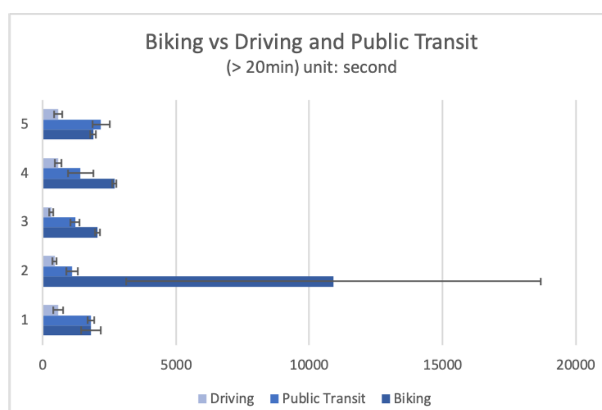


Fig. 16: Transit mode efficacy comparison when bike time is over 20 minutes

B. Discovering Popular Routes

From our visualization of common routes, we find that during morning hours, the most popular routes are around MIT, South Station and Nashua Street at Red Auerbach Way (near TD Garden). MIT as well as Kendall Square are especially big centers for Bluebikes trips. Our interpretation for these observations is that the majority of the bike users use the bikes for commuting purposes - go to school or work with MIT students and employees around Kendall Square, South Station and TD Garden being major users.

During evening hours, we still see that MIT, South Station and Nashua Street at Red Auerbach Way (near TD Garden) are important nodes with popular routes originating or ending at them, but the strengths of common routes vary slightly from the morning common routes.

It is worth noting that at the upper left corner of both maps, there is a strong link between Davis square and Linear park during both morning and evening rush hours. This strong connection is not connected with other nodes where most trips occur. We noticed that the Bluebikes dock at Linear Park is far from either Alewife or Davis Square stations on the Red Line, while the Bluebikes dock at Davis Square is very close to Davis Square station. That is to say, for people living close to Linear Park, it is not convenient for them to ride the Red Line if they choose to walk to subway stations. However, it will be convenient and accessible for them to reach Davis Square station by using Bluebikes.

From the common route plots for each month, we can see a seasonal pattern which has also been demonstrated in previous sections. We can see that the trips between popular stations such as MIT Stata Center, Kendall Square, Nashua Street and South Station happen no matter the weather conditions, while trips happen between less popular stations reduce during winter times. Moreover, when comparing the morning and evening common routes, we find that in June, many trips start and end at the same Bluebikes station during evening times. This may indicate that

To understand which stations are important in the network, we calculate the eigenvector centrality for all the nodes in the network for the whole study period. We found that during the morning, MIT, MIT Stata Center, Kendall Square, MIT Vassar Street, MIT Pacific Street at Purrington Street are the only nodes that ranked with top 5 eigenvector centrality. This implies active interaction between these stations, which could be MIT students or faculties go to school. During the evening

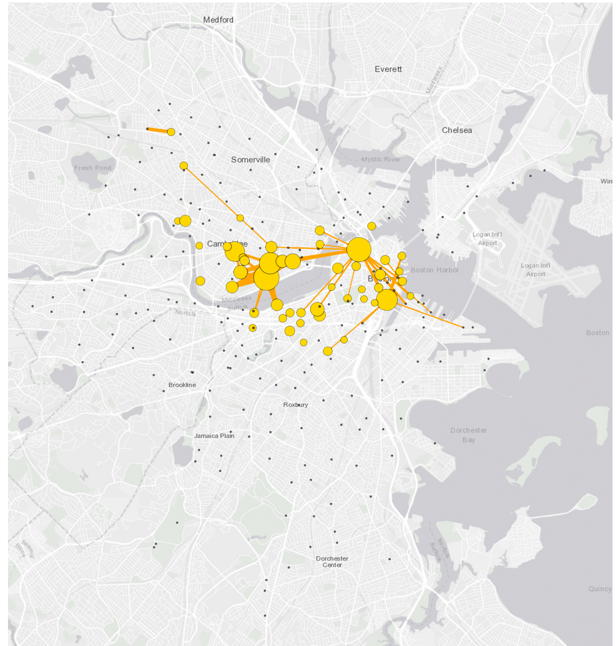


Fig. 17: Common routes from 2017 October to 2018 September in the morning time (07:00-11:00AM).

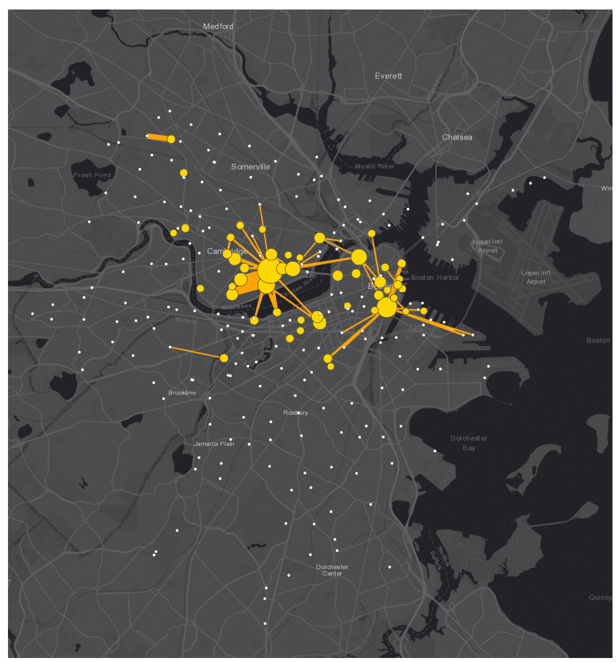


Fig. 18: Common routes from 2017 October to 2018 September in the evening time (15:00-19:00PM).

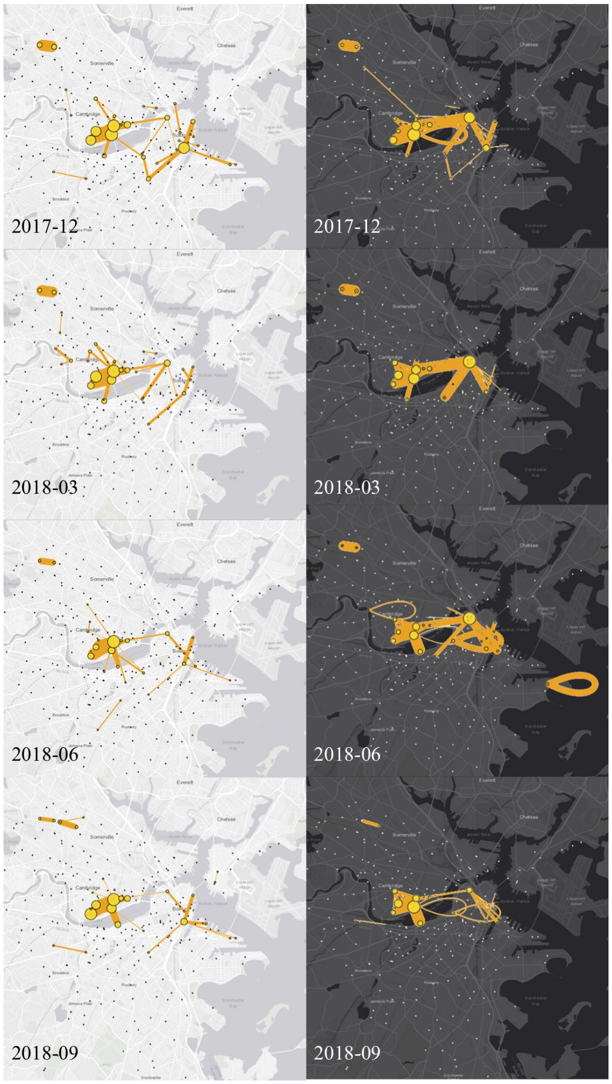


Fig. 19: Common routes for December, 2017, 2018 March, 2018 June and 2018 September. Plots with light maps are popular routes in the mornings; plots with dark maps are popular routes in the evenings.

we see more stations including Nashua Street at Red Auerbach Way, Central Square and Beacon Street at Massachusetts Ave, which seem to be business centers and entertainment centers.

C. Clustering Stations By Usage

We wanted to understand if Bluebikes docks across the city displayed similar usage patterns. In the methodology section, we explained how we performed spectral clustering to obtain 3 clusters of docks based on their usage and location attributes. In this process, each Bluebikes dock was assigned to one of the three clusters, and the geographically distribution of the docks in each cluster can be seen on the maps in Figure 20c and Figure 20a for morning and evening respectively.

By performing clustering on the Bluebikes docks, we were able to identify major transit routes and differentiate between economic hubs and residential neighborhoods in Boston without any information about the city's layout or neighborhood analysis.

According to the 2014 edition of MBTA Ridership and Service Statistics (The Blue Book), the Red Line and the Green Line on the subway system are the mostly popular lines of the subway system [7]. Compare the location of the red cluster (group 1) of both maps with the subway map in Figure 20b. The red cluster (group 1) on both maps follows the Red Line through Cambridge and the Green Line through central Boston (before the lines split in the western part of the city). The Blue Line (traveling northeast of the city towards Winthrop) is the least popular line according to "The Blue Book", and is classified by the spectral clustering with other periphery docks from more residential areas in the third group, shown in green in Figure 20a and purple in Figure 20b.

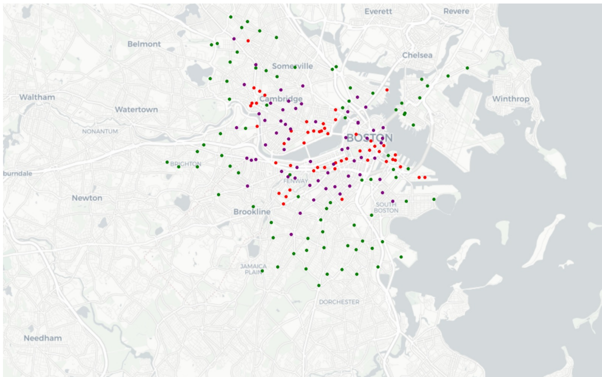
From this classification, we can infer that Bluebikes docks around transit stations are used in similar ways and connect Bluebikes usage to general transit and commuting patterns in Boston. Bluebikes' subscribers use the bike sharing platform to gain access to public transportation.

VII. Applications

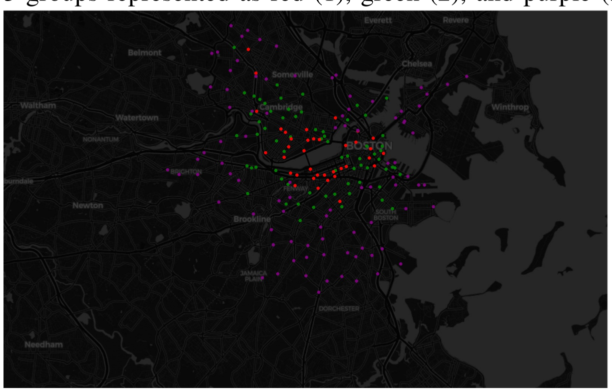
In the previous section, we explored how the Bluebikes system has been used instead of or as an augmentation of the existing public transit system. For mid-distance trips (10-20 minute bike ride), using the bike sharing system is faster than using public transit for popular bike routes. However, users largely reap the benefits of the bike sharing system in the summer and ridership decreases in the winter as seen in Figure 2. We suggest ways to improve the existing MBTA public transit system and based on the findings in this paper to improve accessibility and transit times for users during winter months when bike sharing is less of an option.

We consider the common routes from Figure 17 and Figure 19 and recommend changes to MBTA bus routes. Buses are easier and cheaper to deploy than trains because they do not require heavy infrastructure investment to implement.

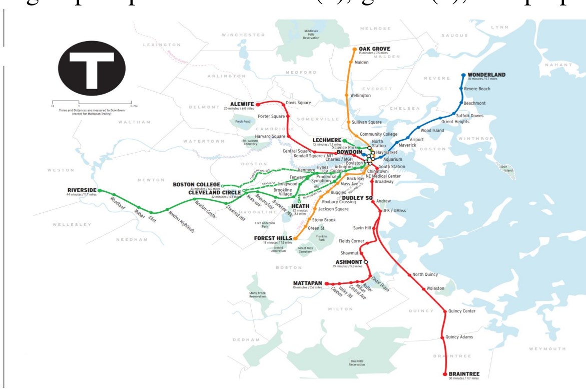
Five of the most popular routes link locations on MIT's campus (either the 67 Massachusetts Avenue or the Stata Center) to locations in Back Bay. Currently the only way to travel these routes by public



(a) The map shows the clustering of the 195 Bluebikes stations based on usage attributes in the morning commute window in 3 groups represented as red (1), green (2), and purple (3).



(b) The map shows the clustering of the 195 Bluebikes stations based on usage attributes in the evening commute window in 3 groups represented as red (1), green (2), and purple (3).



(c) A geographically-accurate map of the MBTA T subway stations in urban Boston provided by Radical Cartography [6]

Fig. 20: A comparison between the geographically positions of our cluster results and the existing MBTA lines.

transportation is using the 1 Bus, which gets stuck in traffic along the Harvard Bridge crossing the Charles River into Cambridge. We recommend the addition of a bus-only lane on Massachusetts Avenue between Hynes Convention Center stop on the Green Line and 77 Massachusetts Avenue in Cambridge. We also recommend more frequent buses on the 1 Bus Route during peak commute times.

There is also a large amount of connections from the Lechmere Station and the North Station stops on the MBTA Green line to the Kendall Square and South Station stops on the Red Line. Currently, the Red and Green lines only connect at Park Street, which does not allow easy transfer between the two most popular train lines at other points in the city. We recommend that a new bus route, potentially an extension of the Silver Line, makes stops at the Bluebikes docks within central Boston with the highest degree (South Station, Haymarket, North Station, Lechmere, and Kendall Square) to connect the Red and Green Lines.

VIII. Conclusion

The Bluebikes system has been adopted by many Boston users for commuting, whether for full end to end trips or as first-mile last-mile supplements to connect them to existing transportation systems including the T. In addition to providing another commute option, Bluebikes provides the added benefit of saving time along popular mid-distance routes in comparison to taking public transit. We hope the city of Boston can apply the knowledge about Bluebikes system usage in order to supplement the existing public transportation infrastructure along popular Bluebikes routes during winter months. This way, when users are less inclined to expose themselves to the elements, residents of Boston still have enough options for their commutes.

IX. Appendix

All code supporting our methodology and analysis can be found on our github: https://github.com/meiaalsup/ids131.

References

[1] Motivate International, Inc. Bluebikes System Data. Bluebikes Boston, www.Bluebikes.com/system-data.
[2] Kabra et. al. "Bike-Share Systems: Accessibility and Availability"
[3] Fishman, Elliot. "Bikeshare: A Review of Recent Literature"
[4] Gebhart, Kyle, Robert Noland. "The impact of weather conditions on bikeshare trips in Washington, DC"
[5] Martin, Elliot and Susan Shaheen. "Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities"

[6] Rankin, Bill, and Louis Hyman. THE BOSTON T. Radical Cartography, 2003, www.radicalcartography.net/?bostonnow.
[7] MBTA Ridership and Service Statistics (The Blue Book). Massachusetts Bay Transit Authority, July 2014.