TAKING THE TRAIN?

Explaining MBTA Commuter Rail Ridership

INTRODUCTION

The MBTA Commuter Rail provides service from suburbs in the Boston Metro Area to Boston area stations, with terminal stations at North Station and South Station. While using commuter rail may be faster, particularly at rush hour, than using a personal vehicle or other transit alternatives, people still choose not to use the Commuter Rail, as can be demonstrated by the high volume of people driving at rush hour.

This study seeks to understand the personal vehicle and public transit alternatives to the MBTA Commuter Rail at each stop to understand what options people have when deciding to use the Commuter Rail over another mode and what characteristics of the alternatives may inspire people to choose them over Commuter Rail. Understanding what transit and driving alternatives are like at each Commuter Rail stop may offer insight into why people are choosing or not choosing Commuter Rail for their trips to Boston, and how to encourage ridership. This project explores what transit and driving options exist for people at every Commuter Rail station during weekday peak periods and how these alternatives, along with other commuter rail factors, might impact the number of Commuter Rail riders at a station level.

Other studies have examined what factors predict and explain transit ridership through multivariate regressions. Vehicle capacity, employment density, service level and integrated ticketing have been found to affect ridership on bus rapid transit, light rail and street cars (Currie et. al 2013). A study on peak Metro ridership in Montreal found that average income, bus service connectivity, distance to a central station and service frequency influenced ridership at the station level (Chan and Miranda-Moreno 2013). Seasonal change and fares also have been found to impact bus ridership specifically (Chiang et al. 2011).

METHODS

The methodology of this paper was derived from other studies that have examined ridership through multivariate linear regressions, however, this project focused on commuter rail ridership. The dependent variable for this analysis used 2012 commuter rail inbound boardings collected by CTPS. The independent variables for this analysis were created and collected using spatial and non-spatial data.

The environmental variables included in this analysis were median income within a half mile, population density within a half mile and job density within a half mile. These factors were calculated using walk buffers generated based on ESRI's network dataset on ArcGIS Pro.

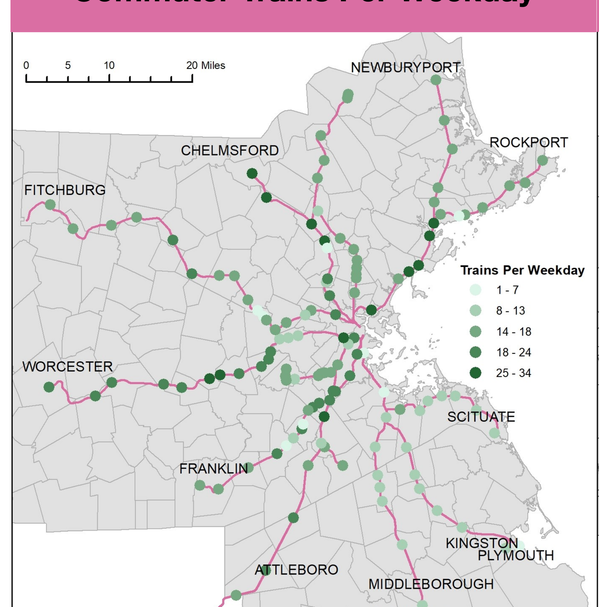
The variables for commuter rail service included the 8AM commuter rail time to the Boston terminal station, the distance as the crow flies from a commuter rail stop to the Boston terminal station, the number of trains going inbound to the Boston terminal station on a weekday and commuter rail on time performance, calculated by using Google's transit directions, point distance to Boston stations, commuter rail schedules and the MBTA dashboard, respectively.

The study also examined driving time to the terminal station on a peak weekday using ArcGIS Pro network analyst, the average number of transit trips per day at transit stops within a half mile buffer of commuter rail stations, transit time to Boston at weekday peak, and the distance from a commuter rail station to an MBTA rapid transit station, using point distance.

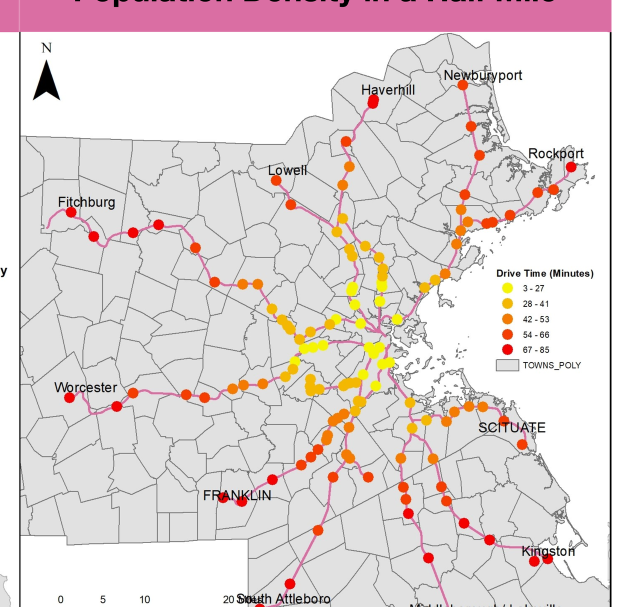
After all dependent and independent variables were collected and cleaned, the distribution of the variables were graphed. The natural log was taken of variables, including jobs and population within a half mile, to normalize the distribution. Once the variables had been normalized, preliminary regressions were conducted to understand the multicollinearity between variables. After multicollinearity was demonstrated, a bivariate correlation analysis was conducted on all commuter rail variables with demonstrated multicollinearity. Single variable linear regressions were then conducted on all correlated variables to predict inbound boardings, to find which of them were most predictive.

Three final regressions were conducted with finalized variable: a regression on environmental and commuter rail variables, a regression on transit variables and a regression using the most predictive variables from the first two regressions. As not all commuter rail stations have transit, the second two regressions had a lower sample size.

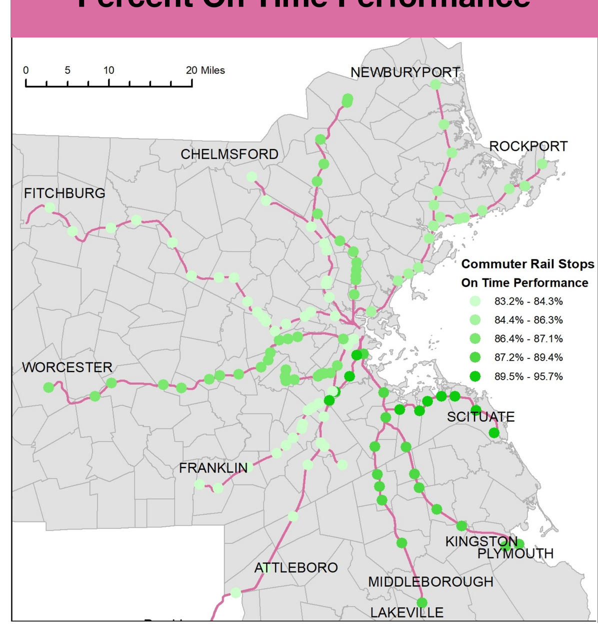
Commuter Trains Per Weekday



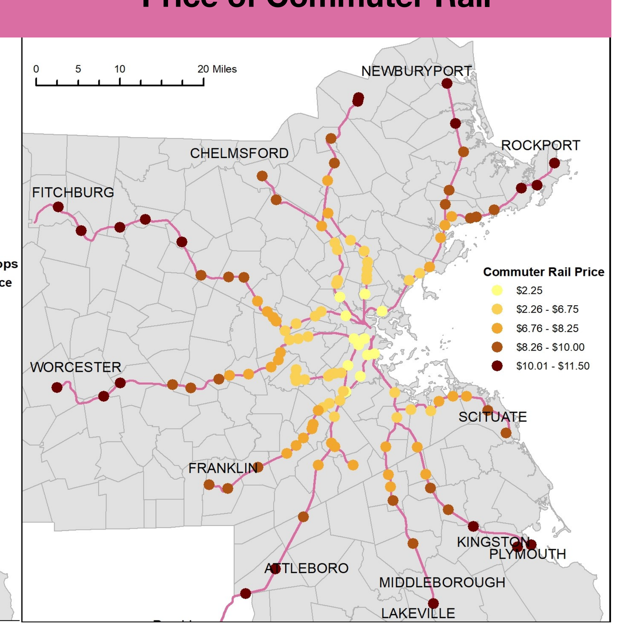
Population Density in a Half Mile



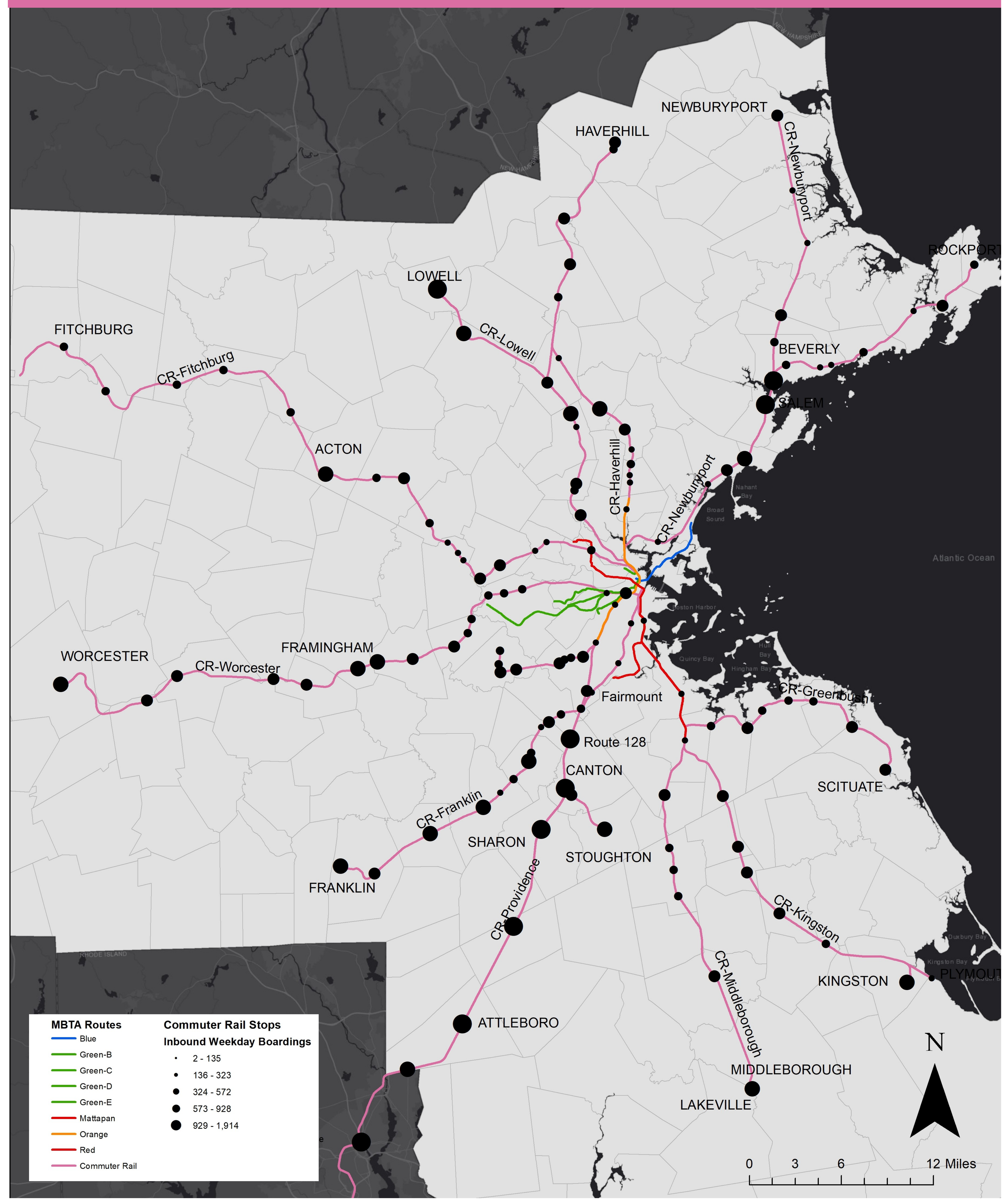
Percent On Time Performance

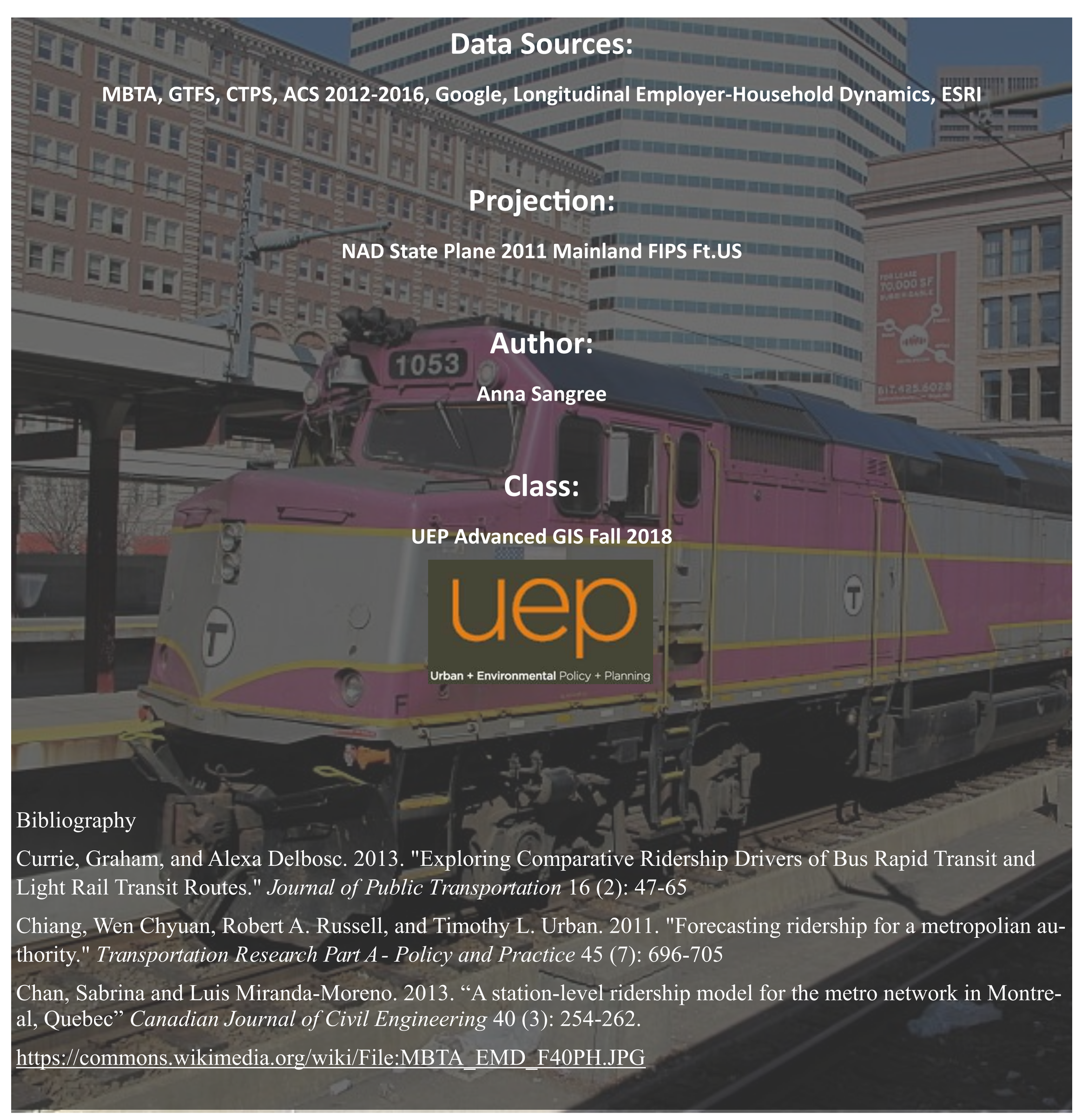


Price of Commuter Rail



RIDERSHIP BY STATION





Data Sources:

MBTA, GTFS, CTPS, ACS 2012-2016, Google, Longitudinal Employer-Household Dynamics, ESRI

Projection:

NAD State Plane 2011 Mainland FIPS Ft.US

Author:

Anna Sangree

Class:

UEP Advanced GIS Fall 2018

Bibliography

Currie, Graham, and Alexa Delbosc. 2013. "Exploring Comparative Ridership Drivers of Bus Rapid Transit and Light Rail Transit Routes." *Journal of Public Transportation* 16 (2): 47-65

Chiang, Wen Chyuan, Robert A. Russell, and Timothy L. Urban. 2011. "Forecasting ridership for a metropolian authority." *Transportation Research Part A - Policy and Practice* 45 (7): 696-705

Chan, Sabrina and Luis Miranda-Moreno. 2013. "A station-level ridership model for the metro network in Montreal, Quebec" *Canadian Journal of Civil Engineering* 40 (3): 254-262.

https://commons.wikimedia.org/wiki/File:MBTA_EMD_F40PH.JPG

Which Commuter Rail Variables Explain Commuter Rail Ridership?

R	R Square	Adjusted R Square	Std. Error of the Estimate
.704[a]	0.495	0.462	261.540

	Beta	P. Value	VIF
(Constant)		0.153	
Median Income in a Half Mile	-0.004	0.960	1.852
Population in a Half Mile (Ln)	0.056	0.618	2.962
Parking_Dummy	0.037	0.645	1.526
Terminal_Dummy	0.238	0.001	1.179
Jobs within a Half Mile (Ln)	0.054	0.603	2.525
Peak On Time Performance	-0.206	0.009	1.420
Peak Drive Time	0.385	0.000	2.588
Trains Per Weekday	0.536	0.000	1.244

Which Transit Alternative Variables Explain Commuter Rail Ridership?

R	R Square	Adjusted R Square	Std. Error of the Estimate
.639[a]	0.408	0.348	270.336

	Beta	P Value	VIF
(Constant)		0.831	
Distance to Rapid Transit (Miles)	1.024	0.000	4.767
Transit Time (8AM)	-0.614	0.030	6.233
Transfer Required	0.141	0.424	2.516
Transit Price (8AM)	0.328	0.028	1.739
Average Trips Per Day of Transit within a Half Mile	0.101	0.517	1.998

CORRELATED INDEPENDENT VARIABLES

	Distance to Terminal Station	Commuter Rail Price	Commuter Rail Time (8AM)	Transit Time (8AM)	Distance to Rapid Transit (Miles)	Drive Time (8AM)
Distance to Terminal Station	1	.860**	.865**	.888**	.864**	.917**
Commuter Rail Price	.860**	1	.718**	.780**	.700**	.804**
Commuter Rail Time (8AM)	.865**	.718**	1	.837**	.723**	.793**
Transit Time (8AM)	.888**	.780**	.837**	1	.888**	.846**
Distance to Rapid Transit (Miles)	.864**	.700**	.723**	.888**	1	.829**
Drive Time (8AM)	.917**	.804**	.793**	.846**	.829**	1

RESULTS

Commuter Rail Variables

Distance to Boston, distance to rapid transit, price of commuter rail, commuter rail time, transit time, and drive time are all highly correlated. This makes sense as they all essentially measure distance to Boston in dollars, minutes and miles.

For the commuter rail variables analysis, trains per weekday (standardized beta=.536, p=.000), drive time at 8AM (standardized beta=.385, p=.000), peak on time performance (standardized beta=-.206, p=.009) and the terminal station (p=.001) were found to be significant. Interestingly, all variables calculated for the area a half mile from commuter rail stations (population, jobs and median income) were not significant.

Transit Variables

Of transit variables, miles to rapid transit (standardized beta=1.024, p=.000), transit time at 8AM (-.614, p=.030) and the price of transit at 8AM (standardized beta=.328, p=.000) were significant in predicting commuter rail ridership, however, transit at 8AM and miles to rapid transit displayed high multicollinearity. Transfers, and the average trips per day on transit within a half mile were not significant.

Transit and Commuter Rail Variables.

After finding the most significant commuter rail and transit variables, a regression was run on the most significant variables from each regression with an adjusted R2 of .573. In the final regression, trains per day (standardized beta=.463, p=.000), transit time at 8AM (standardized beta=-463, p=.048), and drive time (standardized beta=.443, p=.020) and miles to rapid transit (standardized beta=.640, p=.047) were significant.

CONCLUSION

This study found that the number of trains per day, or frequency of service and the distance to Boston, in the form of price, drive time, and Euclidean distance, have the largest impact on commuter rail ridership. Of distance variables (all very highly correlated) the driving time to Boston at peak is the most predictive of commuter rail ridership. This demonstrates that the more trains the MBTA runs, the more people will ride the commuter rail. However, this may not be a causal relationship, as the MBTA likely runs more trains because there is high ridership. High on-time performance had some impact on higher commuter rail boardings as well, which is expected.

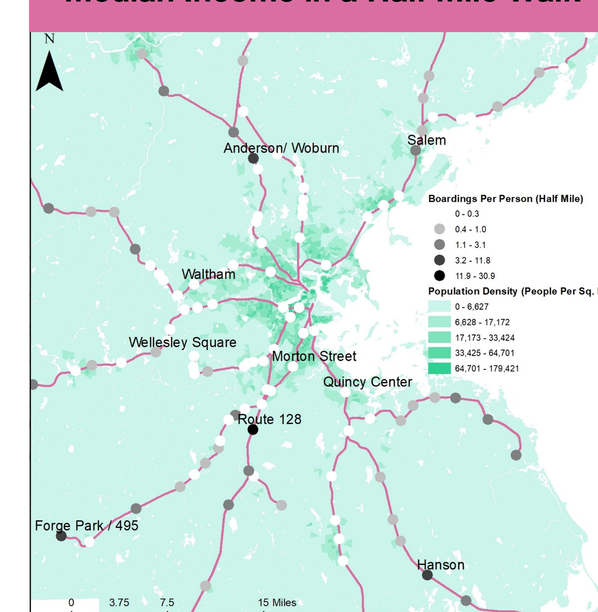
Of transit variables, miles to rapid transit, and transit time to Boston at 8AM are the most predictive of commuter rail ridership, which again demonstrates the relationship between distance from Boston and commuter rail ridership. This may show that places close to Boston, that have transit service or are near a rapid transit station, have lower ridership because people have an alternative to commuter rail. However, as transit time gets longer, ridership on commuter rail increases, demonstrating that, as transit becomes less desirable and more time consuming, people switch to commuter rail. Future research can examine the connection between commuter rail and transit price and times by adding a variable that is the difference between the price and time of commuter rail. Adding a dummy variable that assesses whether having a transit option impacts commuter rail ridership could also be added.

Finally, population and other variables within a half mile are not significant to the commuter rail. This is likely because commuter rail stations often attract people from far away from the station. This makes weighting the ridership by population very difficult, as it is difficult to know where commuter rail riders come from depending on the station. For example, Route 128 has 30 boardings for every one person who lives within a half mile of the commuter rail station.

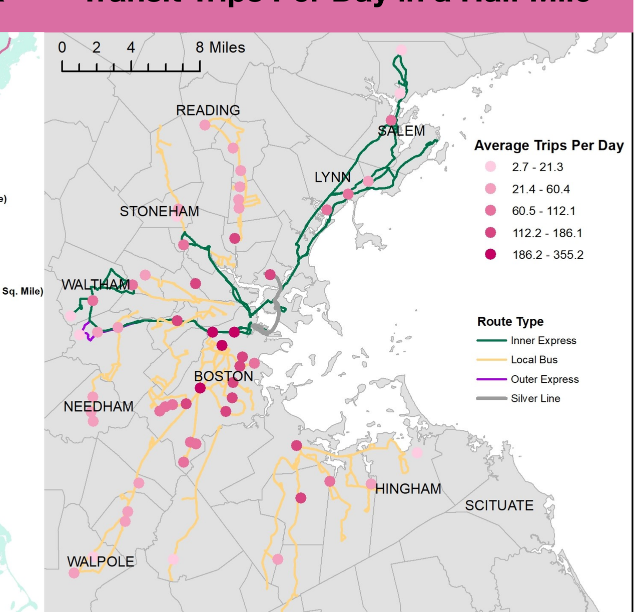
Finally, this analysis had limitations. The data on boardings is from 2012. Also some stops could not participate in the regression because they are new since 2012. Finally, population buffers of a half mile were found to be inadequate. Further analysis should integrate a variety of buffer distances.

This analysis has begun to demonstrate the relationships between transit, driving and commuter rail. These transportation relationships can inform policy aimed at affecting and changing transportation patterns in Massachusetts.

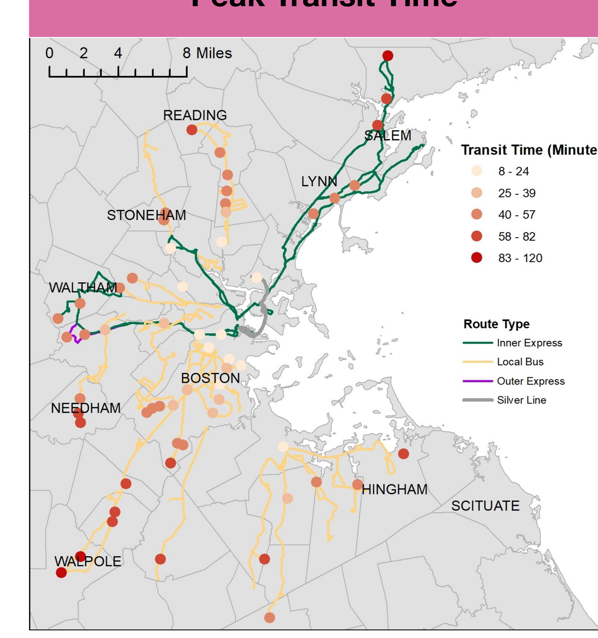
Median Income in a Half Mile Walk



Transit Trips Per Day in a Half Mile



Peak Transit Time



Miles to Rapid Transit