# Factors Affecting MBTA Commuter Rail Ridership

## Developing a Regression Model to Forecast Station Ridership

## Introduction

The MBTA commuter rail system provides more passenger miles per day than the three MBTA heavy rail transit lines combined (Red, Blue, and Orange), yet is rarely discussed compared to these modes. This project aims to isolate the most significant variables determining ridership in order to direct attention to the most effective changes to expand service. With ever growing car traffic throughout the area, improvements to the commuter rail system have the potential to benefit all, not just those who use it every day. By decreasing transit times, everyone benefits from increased free time, not to mention that even with fossil fuel powered trains, emissions per passenger mile are much lower than those of car trips. Finally, the capital cost for much of the system was covered many years ago and future investments could prove to be a very efficient use of limited government funds. Even in its current state with effectively stagnant ridership over the last decade, the commuter rail system manages to achieve a fare box recovery ratio (% operating expenses covered by fare revenue) of 50% which is in line with effective car trip subsidies, not counting the increased externalities of car trips. Considering non-operating revenues, the recovery ratio approaches 72% which is quite high for public transit. In addition, it is an important time to discuss these changes since about half of coaches and 40% of locomotives are beyond their 25 year service lives. The decisions we make today will shape the future of the system going forward a generation, making this a critical time to make significant changes to better use the significant infrastructure we have built over the last century and a half.

## Methodology

Calculating the passengers within different zones was done by first creating network analysis service area polygons. Speed attributes were used to create the drive time service areas, and distance was used for the two walking areas. To get the most accurate population calculation, these areas were then intersected with census blocks, and their sum was used to calculate population in each service area. Trains per day and minimum trip time were calculated from MBTA schedule data. Parking was extracted from MBTA parking data and the Rapid Transit and Terminal stations were geocoded from the rail map.

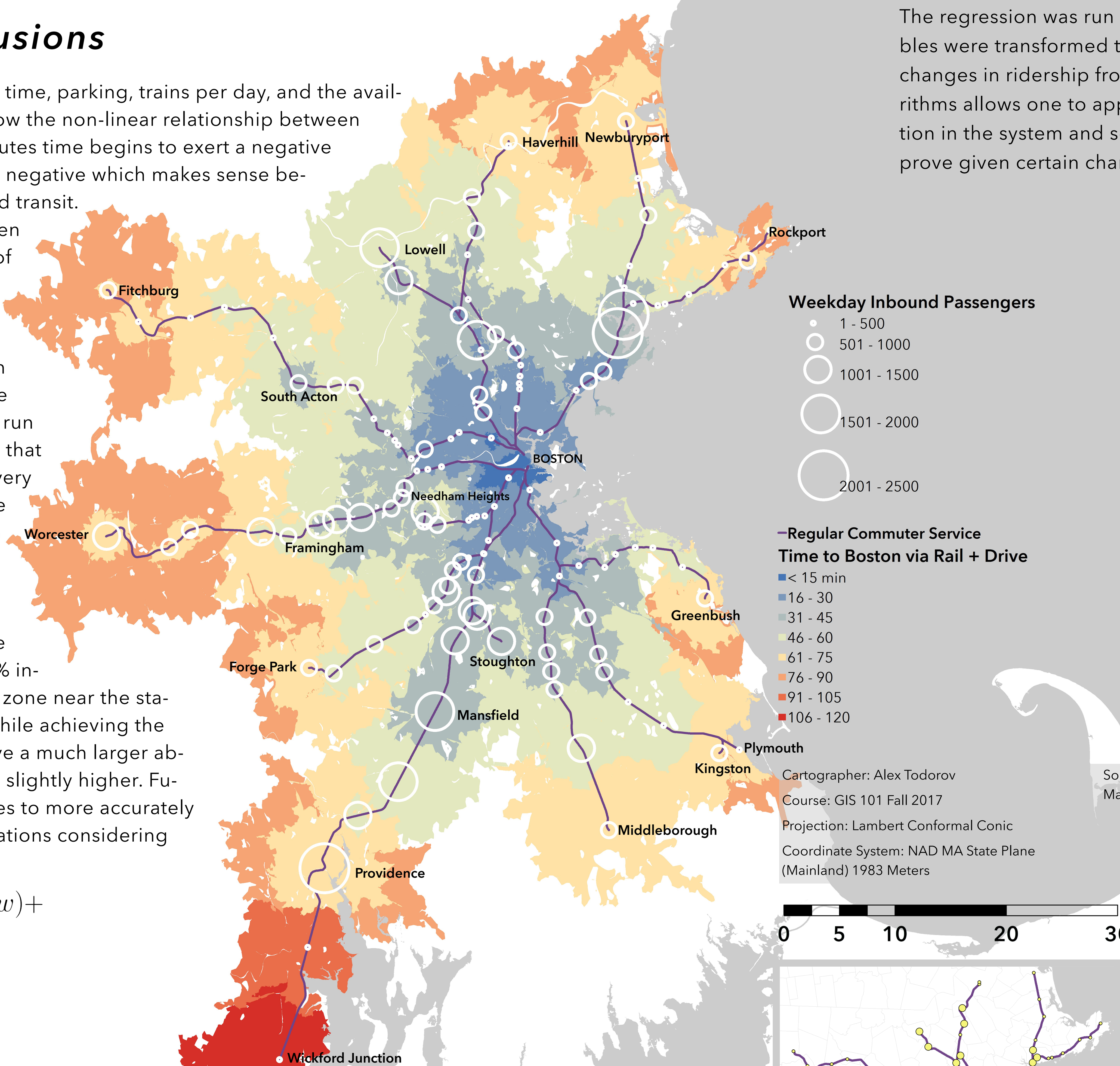
The regression was run in STATA and most of the variables were transformed to logarithms to get the percent changes in ridership from the coefficients. Using logarithms allows one to apply the coefficients to any station in the system and see how its ridership should improve given certain changes.

## Conclusions

The most significant variables were minimum morning trip time, parking, trains per day, and the availability of rapid transit. Minimum trip time is squared to show the non-linear relationship between time and ridership. The coefficients show that after 50 minutes time begins to exert a negative impact on passenger numbers. The sign for rapid transit is negative which makes sense because stations closer to Boston are the only ones with rapid transit. The frequency of rapid transit makes it more appealing even though the commuter rail may be slightly faster. The sign of trains per day is also positive which is reasonable since higher frequencies make rail a more convenient choice. There are concerns about endogeneity since it is likely that lines with high numbers caused the service increase in the first place. In addition, analyzing a decrease in trip time could result in an increase in frequency since the train can run more trips per day, so when analyzing these two variables, that relationship should be taken into account. Parking is also very significant, but there are endogeneity concerns since more used stations would typically have more parking spots. On their own, the population numbers were not significant but combined they are. As expected, the coefficient for 5-minute walks was higher than that of the 10-minute walks. With the log transformation this means that a 10% increase population of the area around the station results in a 1.15% increase in boarding. A 1% increase in ridership in the small zone near the station could only involve a variance for one or two parcels while achieving the same increase from the 10-minute drive zone would involve a much larger absolute change in population even though the coefficient is slightly higher. Future research could involve more granular population zones to more accurately model the decay in demand and use more detailed calculations considering traffic and turn delay to generate service areas.

$$ln(passengers) = .35ln(parking) + .11ln(pop5w) + .027ln(pop10w) + .029ln(pop5d) + .16ln(pop10d) + 1.08ln(trainsperday) + .09term - 1.97rapid + .070mintime - .00067mintime^2 - 3.09$$

| *Variable* | *Coefficient* | *P>\|t\|* |
|---|---|---|
| ln(Station Parking) | 0.3515 | 0 |
| ln(Population 5 min walk) | 0.1145 | 0.089 |
| ln(Population 10 min walk) | 0.0276 | 0.678 |
| ln(Population 5 min drive) | 0.0289 | 0.789 |
| ln(Population 10 min drive) | 0.1569 | 0.18 |
| ln(Trains per Day) | 1.0762 | 0 |
| Terminal Station | 0.0943 | 0.51 |
| Rapid Transit Available | -1.9733 | 0 |
| Minimum Morning Trip Time | 0.0699 | 0 |
| Min. Morning Trip Time$^2$ | -0.0007 | 0 |
| Constant | -3.0966 | 0.01 |
| R-Squared | 0.82 | |
| N | 112 | |
| SER | 0.43 | |




Cartographer: Alex Todorov
Course: GIS 101 Fall 2017
Projection: Lambert Conformal Conic
Coordinate System: NAD MA State Plane (Mainland) 1983 Meters

Sources: American Community Survey, ESRI, MassGIS, MBTA