Blue Bike Membership vs. Location

A Look Into the Relationship Between Blue Bike Membership and Space

Vaibhavi James, Wellesley College c.o. 2025, Data Science Capstone

Introduction

- Riders in different parts of greater Boston may have Blue Bike memberships depending on how frequently they use Blue Bike
- Hypothesize areas with more popular stations will have a higher proportion of riders with a Blue Bike membership
- Interested in whether location can be an indicator of higher proportions of Blue Bike members having a membership

Research Question

Does location play a role in whether or not Blue Bike users have a membership?

Data & Methods

Primary data sources

- Blue Bike trips: start/end stations (ID, name, coordinates), membership status
- Blue Bike stations: station name, coordinates, ID, seasonal status, municipality

Data manipulation

- Separate datasets for ride information and station information were combined to form a dataset with detailed station information for each trip
- Analysis was limited to Boston and Cambridge since the data for other municipalities was very limited
  - More likely to see significant spatial patterns in these areas since that's where riders are concentrated

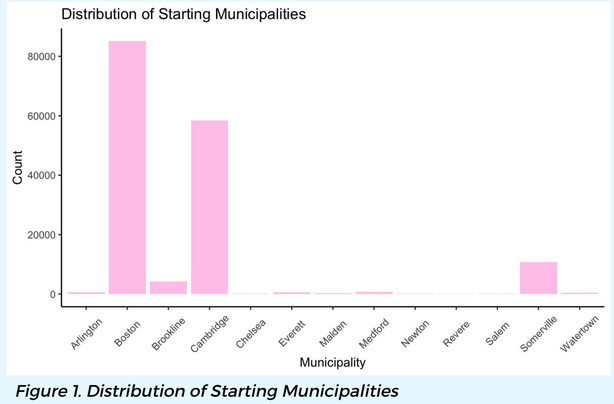

Figure 1. Distribution of Starting Municipalities

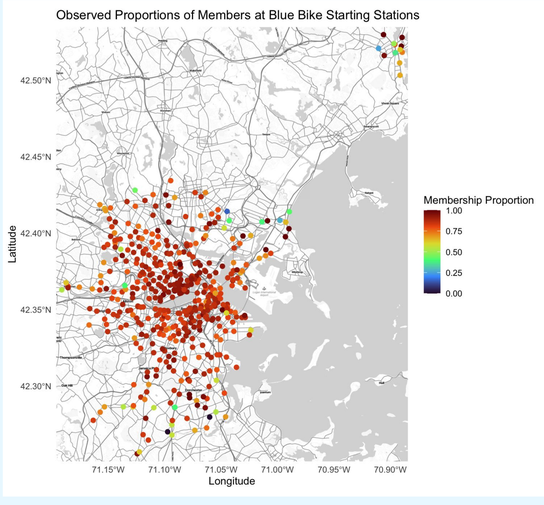

Figure 2. Observed Starting Member Proportions

Results

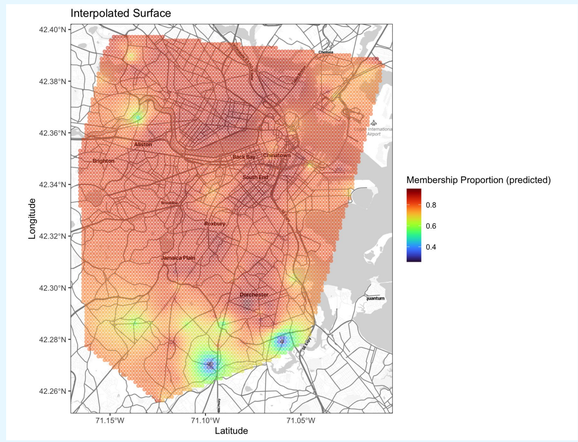

Figure 3. Predicted Membership Proportions in Boston and Cambridge

Model

- Total rides and number of docks used as additional covariates
- Assumptions:
  - Linearity of covariates
  - Stationarity, independence, zero mean of spatial residuals
- Variogram fitted automatically
- Validated using k-fold validation

$$\widehat{\text{Membership Proportion}}(s_0) = \sum_{i=1}^{n} \lambda_i [\beta_0 + \beta_1 \text{Total Rides}(s_i) + \beta_2 \text{Docks}(s_i)]$$

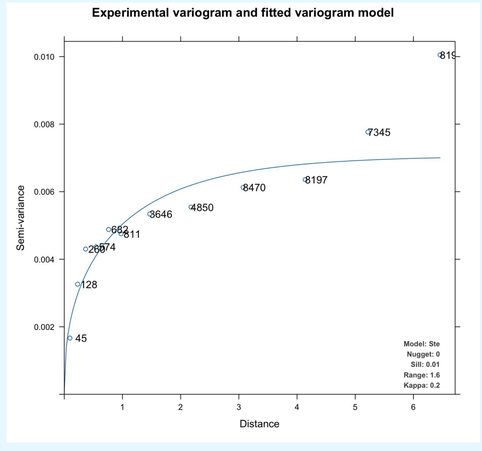

Figure 4. Sample and fitted variogram model.

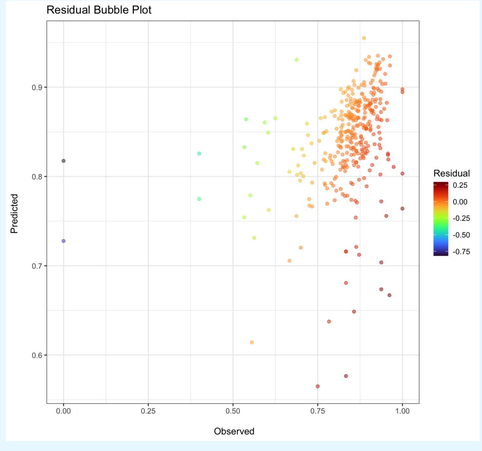

Figure 5. Bubble plot of residuals.

Conclusion

- The **predicted membership proportions across Boston and Cambridge appear to be uniform**
  - This could indicate location is not a strong factor in whether or not riders starting at a given station have a membership
- Model is limited by only looking at how many members begin their trip at each station, does not take into account destinations/ending stations

Future directions

- Include more detailed demographic information for each station in the model

For the code and data, see my GitHub repository