# Is Bikeshare Really Equally Accessible and Affordable to Everyone in Boston?

## Background

- In a previous project (regression and city exploration 3), I explored the relationship between **income and bikeshare usage in Washington, D.C.**
- The findings showed that **low- to mid-income populations tend to ride bikes more as income increases up to a point**. However, they **still face physical and financial barriers** to access.
- **The City of Boston** has been proactively working to expand affordability and accessibility through **new infrastructure and a discounted membership program (launched in January 2018).**

## Objective

- **To examine how income affects bikeshare usage by comparing 2015 and 2019 Bluebike data** (before and after the implementation of the discount membership).

# Data Preparation

✓ **Clean the record-level data and merge them by GEOID or GISJOIN to create census tract-level datasets for income-based bike usage analysis.**

## Packages Used

- ggplot2
- dplyr
- readr
- Sf
- tmap
- scales

## Bluebike Data 2015, 2019 (record)

- Load Bluebike 2015, 2019 *bind_rows(), read_csv()*
- Summarized **total bike usage** *total_users* by start station name *group_by(), summarise ()*
- Converted into an sf object and joined with the tract shapefile *st_transform(), st_join()*

## ACS Data 2011 – 2015, 2015 – 2019 (census tract)

- Loaded 5-year ACS data and shapefiles including household income, income-to-poverty ratio, commute time to work, and race variables *read_csv()*
- Renamed variables with more intuitive names using the NHGIS codebook *rename()*

## H+T Index Data 2016, 2022 (census tract)

- Loaded the H+T Index 2016 and 2022 *read_csv()*
- Renamed relevant columns for consistency in R *rename()*

## Merged Datasets (census tract)

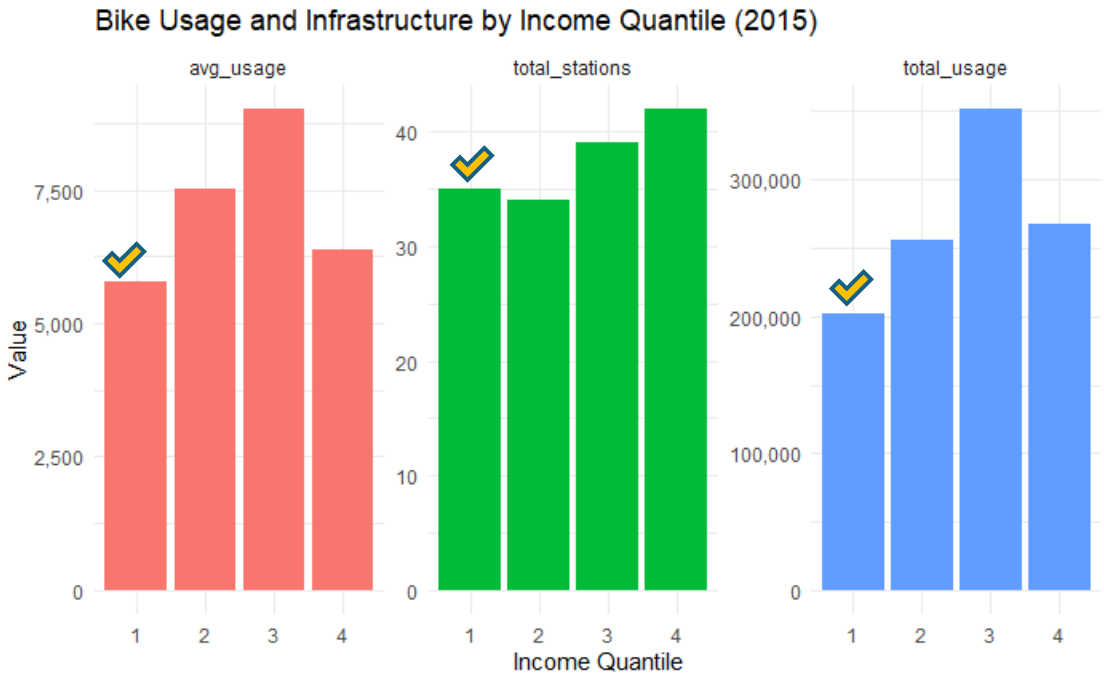
- Merged all datasets using GEOID/GISJOIN *merge(), join()*
- Created new census tract-level variables *mutate(), summarise()*
- **income_quantile (based on median income)** *ntile():*
- **total_stations and total_usage** *sum()*
- **n_tracts** *n()*
- **avg_transport_cost and avg_income** *mean()*
- **avg_usage (total usage divided by number of stations)**

# Analysis - 2015

✓ **Total number of 156 stations and total trip volume 1,107,442 users**

Descriptive statistics generated using *summarise, mean, sum, mutate, filter, group_by, geom_col, facet_wrap* functions

## >Income Quantiles and Bluebike Usage 2015 (census tract level)



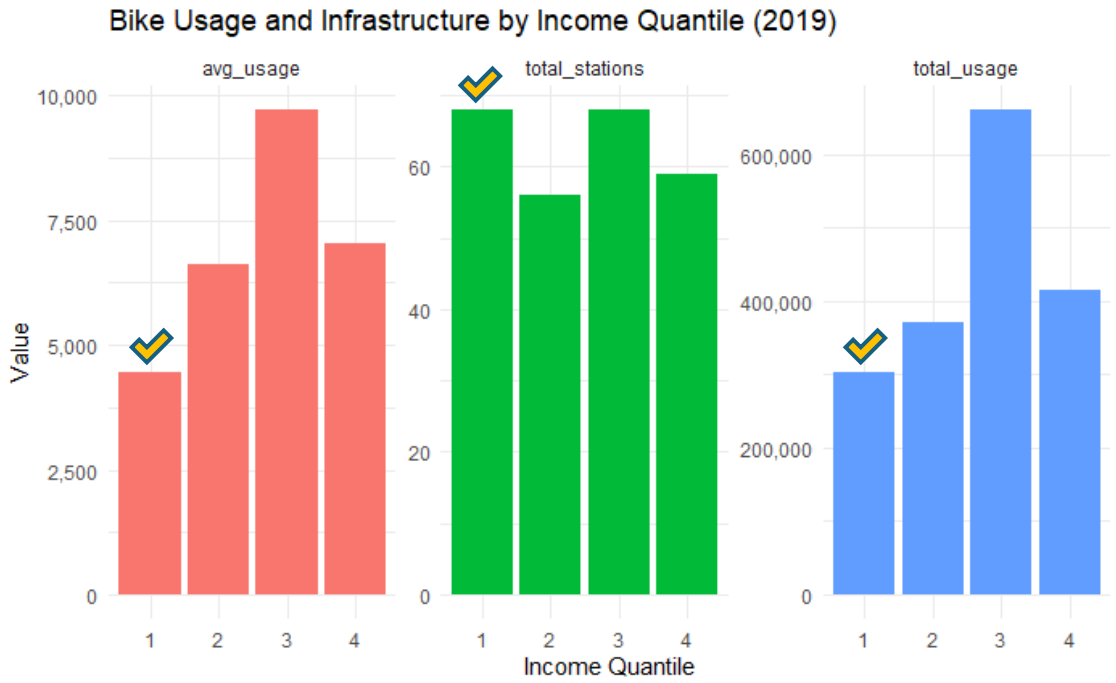

- Wealthier neighborhoods tend to have more **Blue Bike stations**.
- **3rd income quantile** records the **highest total trips (351,859)** and **highest average usage by station (9,022).**
- **1st income quantile** shows the **lowest total trips (202,604**) and **lowest average usage per station (5,788).**
- Despite large income differences, **average transportation cost remains relatively similar** across all quantiles.

| Income Quantile | Average Income* | Average Transport Cost** | Total Stations***(A) | Number of Tracts | Total Usage (B) | Average Usage per Station (B/A) |
|---|---|---|---|---|---|---|
| 1 | $ 29,083 | $ 7,411 | 35 | 23 | 202,604 | 5,788 |
| 2 | $ 57,049 | $ 8,481 | 34 | 23 | 255,801 | 7,523 |
| 3 | $ 82,283 | $ 7,869 | 39 | 23 | 351,859 | 9,022 |
| 4 | $ 112,321 | $ 8,436 | 42 | 22 | 268,109 | 6,383 |

# Analysis - 2019

✓ **Total number of 411 stations and total trip volume 2,522,771 users**

## >Income Quantiles and Bluebike Usage 2019 (census tract level)



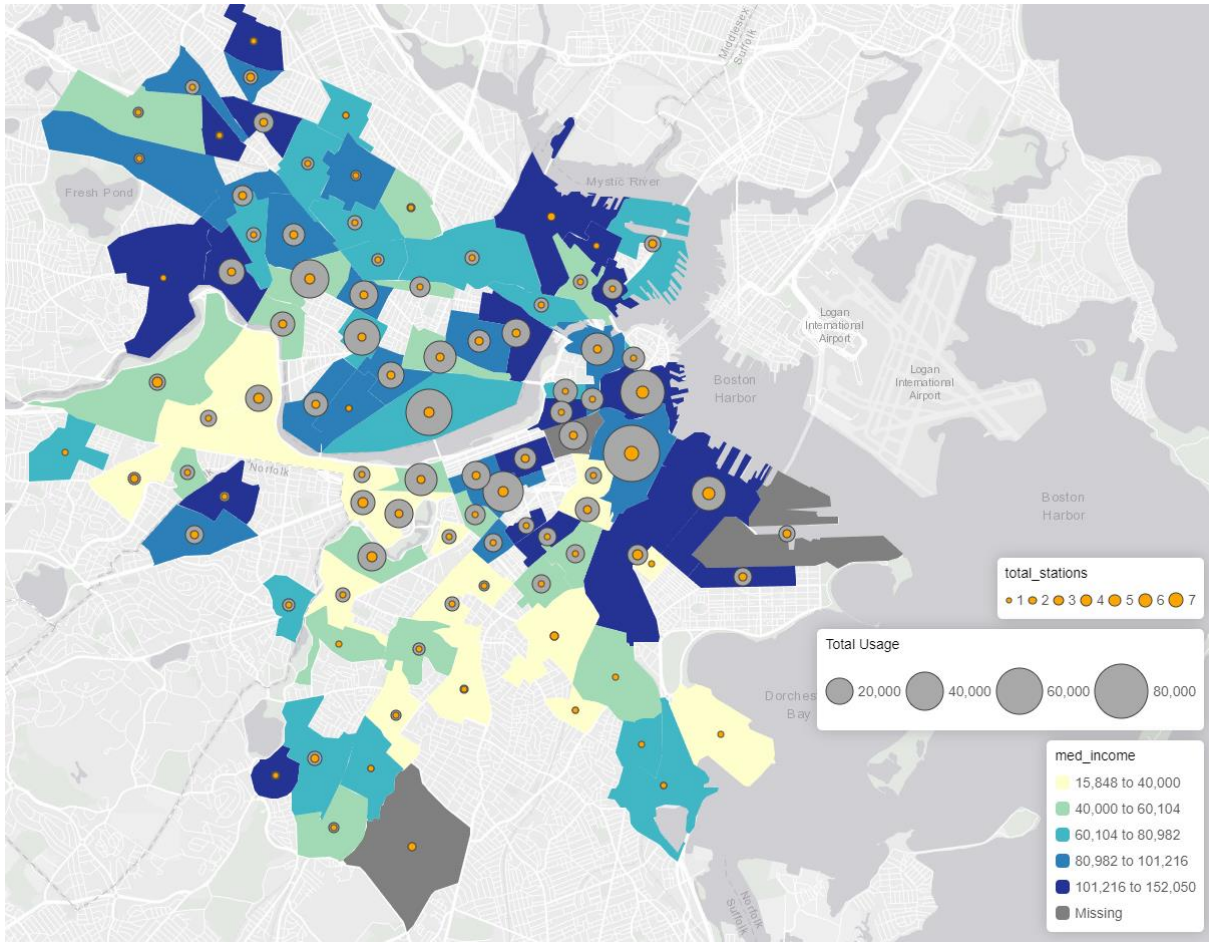

- **1st (lowest) and 3rd (middle**) income quantiles had the **highest number (68) of stations installed.**
- **1st income quantile's total bike usage increased** from 202,604 trips in 2015 to 303,821 trips in 2019.
- Despite the increased infrastructure, **the lowest-income areas still recorded the lowest total and average usage** among all quantiles

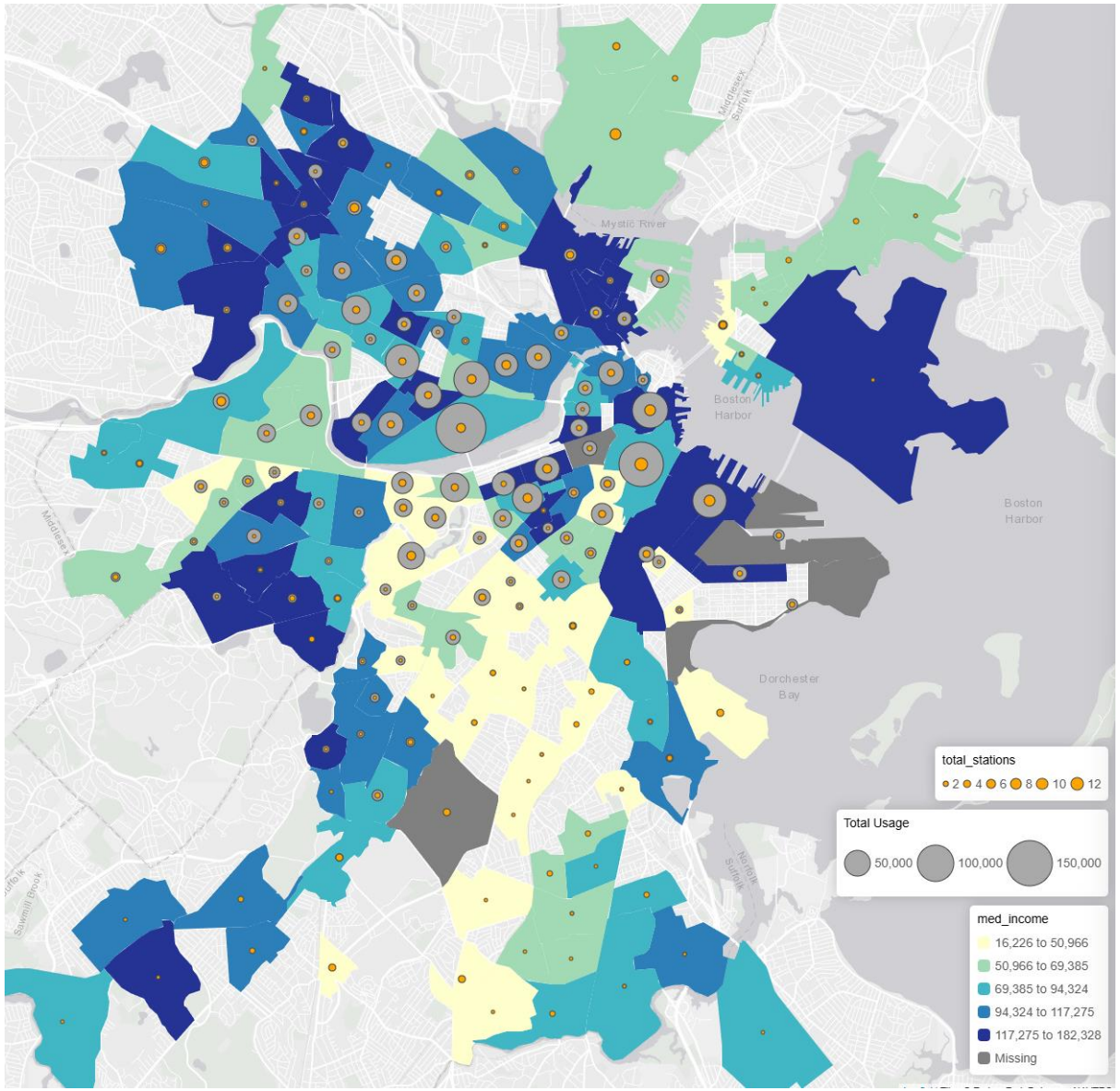
| Income Quantile | Average Income* | Average Transport Cost** | Total Stations*** (A) | Number of Tracts | Total Usage (B) | Average Usage per Station (B/A) |
|---|---|---|---|---|---|---|
| 1 | $ 36,597 | $ 9,766 | 68 | 34 | 303,821 | 4,467 |
| 2 | $ 69,208 | $ 10,234 | 56 | 33 | 369,991 | 6,606 |
| 3 | $ 97,455 | $ 10,207 | 68 | 33 | 661,299 | 9,724 |
| 4 | $ 135,428 | $ 10,526 | 59 | 33 | 414,832 | 7,031 |

# Analysis – 2015, 2019

## >Median Income, Station Count, and Bike Usage 2015, 2019 (census tract level)

*tmap functions (tm_shape, tm_fill, tm_bubbles, tm_layout)*

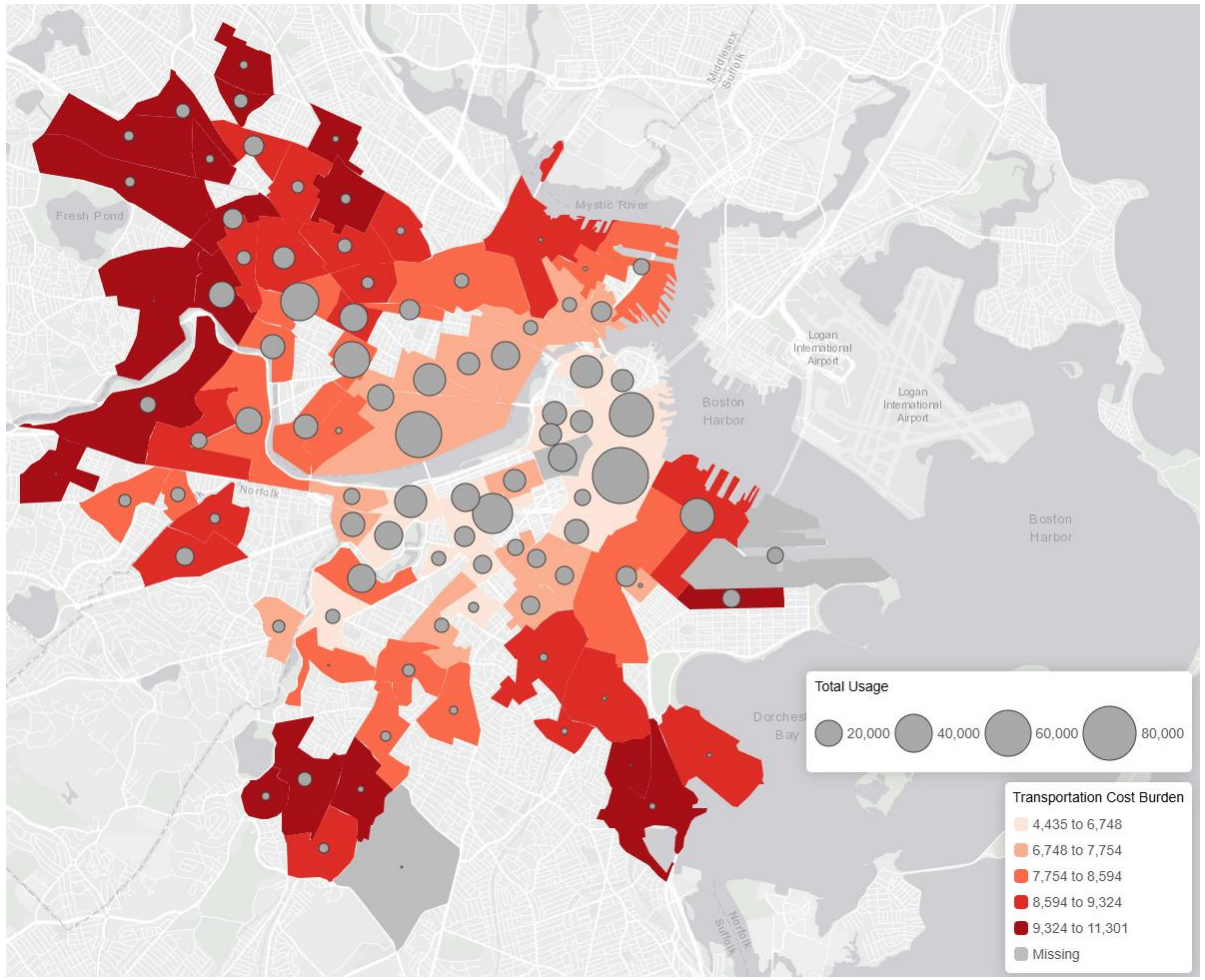


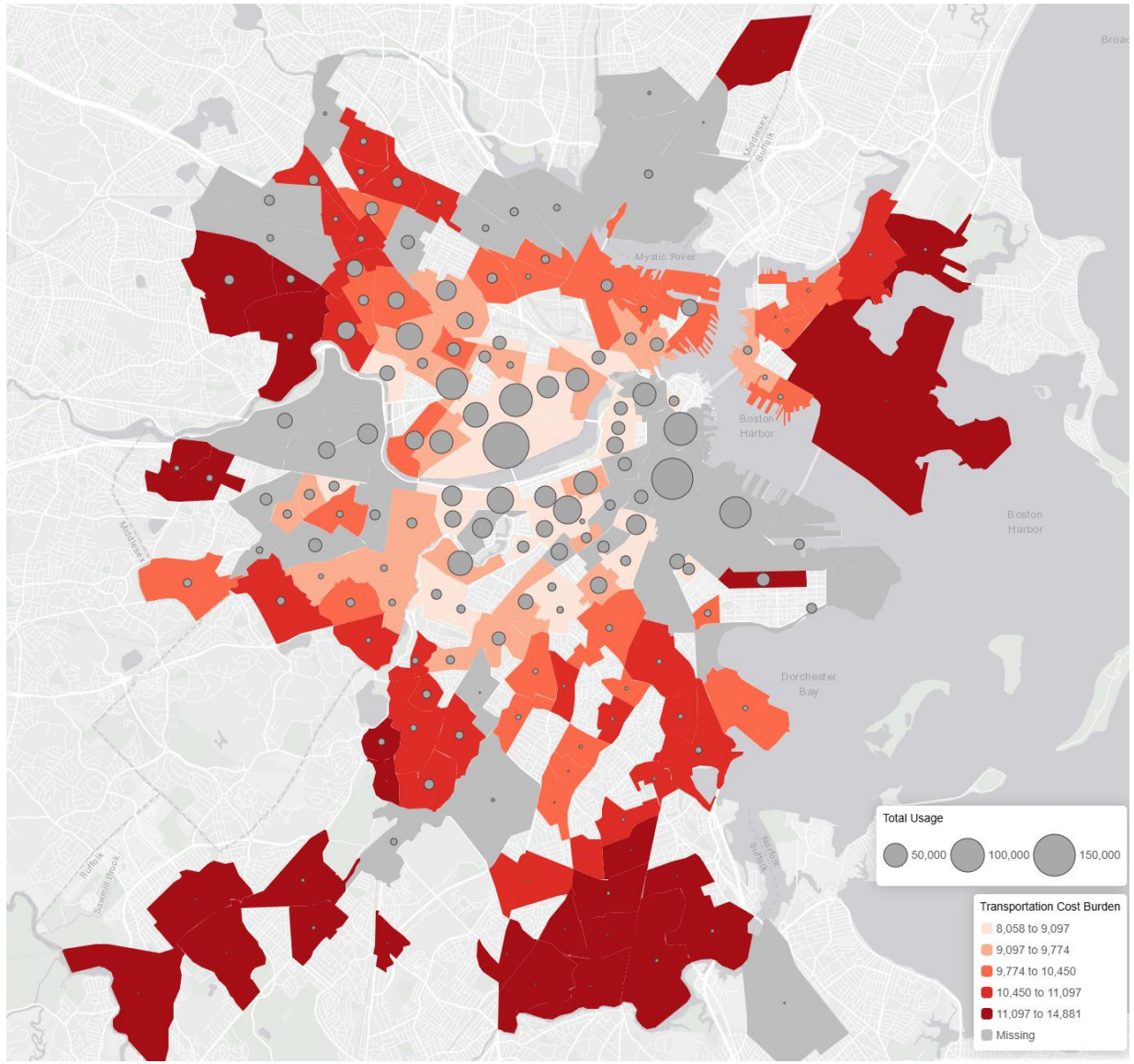



# Analysis – 2015, 2019

## >Transportation Cost and Bike Usage 2015, 2019 (census tract level)

*tmap* functions (*tm_shape, tm_fill, tm_bubbles, tm_layout*)



# Regression Model 2015

- ✓ Regression model was built with *lm* function to understand the relationship between **income and bike usage**.
- ✓ Other socio-economic factors were included to increase the explanatory power of the model.

**>Regression Model 2015**

```
Call:
lm(formula = total_usage ~ med_income + I(med_income^2) + five_to_nine +
    sixty_to_eightynine + t_cost_ami, data = merged_df)

Residuals:
   Min     1Q Median     3Q    Max
-19376  -5783  -1915   4420  46340

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.713e+04  8.392e+03   2.041 0.044374 *
med_income          4.860e-01  1.524e-01   3.188 0.002004 **
I(med_income^2)    -2.489e-06  1.030e-06  -2.416 0.017825 *
five_to_nine        4.714e+01  8.593e+00   5.486 4.15e-07 ***
sixty_to_eightynine -2.272e+01  1.249e+01  -1.819 0.072440 .
t_cost_ami          -3.630e+00  9.102e-01  -3.988 0.000141 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10620 on 85 degrees of freedom
  (결측으로 인하여 3개의 관측치가 삭제되었습니다.)
Multiple R-squared:  0.499,     Adjusted R-squared:  0.4695
F-statistic: 16.93 on 5 and 85 DF,  p-value: 1.39e-11
```

- **Median Income:** Positively associated with bike usage; however, usage declines at higher income levels (inverted U-shape).
- **Commute Time**: Only **5-to-9-minute** category is significant. This means short-distance commuters more likely to use bikes.
- **Transportation Cost: Negative relationship. Tracts with lower transportation costs show higher bike usage.**
- **Race**: **No racial variables were statistically significant** in 2015.

# Regression Model 2019

- ✓ Several changes in the significance of **poverty ratio, race**, and **commute time** compared to 2015
- ✓ Relationship between **household median income and bike usage remains the same**.

## >Regression Model 2019

```
Call:
lm(formula = total_usage ~ med_income + I(med_income^2) + poverty_under0_5 +
    poverty_2_00_over + black + asian + five_to_nine + ten_to_fourteen +
    t_cost_ami, data = merged_df_19)

Residuals:
   Min     1Q Median     3Q    Max
-38920  -9549   -588   6652  91195

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       4.241e+04  1.735e+04   2.445 0.015911 *
med_income        4.960e-01  2.191e-01   2.263 0.025364 *
I(med_income^2)  -2.025e-06  1.058e-06  -1.914 0.057919 .
poverty_under0_5 -2.049e+01  8.325e+00  -2.462 0.015218 *
poverty_2_00_over -6.429e+00  1.842e+00  -3.490 0.000672 ***
black             4.344e+00  1.801e+00   2.412 0.017352 *
asian             1.220e+01  5.577e+00   2.188 0.030585 *
five_to_nine      3.753e+01  1.744e+01   2.152 0.033371 *
ten_to_fourteen   4.698e+01  1.692e+01   2.776 0.006364 **
t_cost_ami       -5.160e+00  1.414e+00  -3.649 0.000387 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16590 on 123 degrees of freedom
  (결측으로 인하여 36개의 관측치가 삭제되었습니다.)
Multiple R-squared:  0.5078,    Adjusted R-squared:  0.4718
F-statistic:  14.1 on 9 and 123 DF,  p-value: 2.039e-15
```

- **Poverty Ratio** became statistically significant (both lowest and highest groups less likely to bike).
- **Commute Time**: **10–14 minute category newly significant**, along with 5–9 minutes.
- **Transportation Cost** **still shows negative relationship** with bike usage.
- **Race**: **Black and Asian populations now show a positive and significant association with bike usage,** possibly due to expanded infrastructure and increased access through discount membership.

# Conclusion

- Boston's 2018 launch of **discounted bikeshare memberships** and the **addition of 250+ new stations improved access**, especially in underserved areas.
- **Stronger effects were expected, but not fully observed**, likely due to the short-term scope of the study.
- **Usage increased significantly in low-income tracts** (in absolute terms) between 2015 and 2019.
- However, **the lowest income quantile still had the lowest average usage.**
- **The relationship between income and bike usage remained similar** in the regression models.
- Still, Boston's efforts to expand access and reduce financial barriers are promising steps toward long-term equity.