Bike Stations in Boston

https://github.com/bluedevil28/JohnsonFromuthCohen_ENV872_EDA_FinalProject

Laurel Cohen, Jeff Fromuth, Blair Johnson

1 Rationale and Research Questions

In the era of the climate crisis, the race is on to find ways to reduce the greenhouse gas (GHG) emissions connected with every possible modern activity and sector of the economy. Cities face unique challenges and opportunities in the realm of sustainability. On a basic level, the major environmental advantage of a city is in population density: urban dwellers tend to live in high numbers close together, instead of enacting the low-density, generally high-carbon-footprint sprawl of the suburbs. In recent years, many cities of various shapes and sizes—from New York City to Dallas—have introduced bike-sharing programs, where users can rent bikes for short periods of time and return them to any one of assorted docking stations around the city. These programs stand to offer a wide range of benefits, from better health for users who increase their exercise to less traffic congestion and air pollution on the roads from fewer cars driven.

As with any other new amenity, it is worth scrutinizing how and where a given municipality with bike sharing chooses to make that amenity available. For our study, we chose to investigate the siting of bike-sharing stations in Boston. The Census data offered thousands of variables, of which we selected several: demographic, including family size and race; financial, including homeownership status and income; and professional, including employment status and mode of commute to work. For this analysis, we sought to determine whether the locations of bike-sharing stations in the city were related to the categories of people found across census tracts. We chose Boston because its data availability was robust and all the members of our group have lived in or near Boston, and we focused on bike sharing programs because several group members are specifically interested in questions of equity and access in the world of transportation. Public officials in Boston have also publicly expressed a desire to make bike sharing available to all, so our team wondered whether that sentiment was evident in the real locations of bike stations in the city.

We chose the following question to anchor our inquiry:
Are bike sharing stations in Boston equally available across population groups?

2 Dataset Information

2.1 Data Retrieval

For this analysis, we used data from two different sources. For population data in Boston, we accessed CSVs from the website of the United States Census, using the data from the 2016-2020 American Community Survey (ACS). The ACS is conducted every year and is used to determine how hundreds of millions of state and federal funds are distributed annually (Census Bureau 2022). For data on the locations of bike sharing stations in Boston, we downloaded shape files from the Blue Bikes Boston website. To pull the Census data into R, we used a package called tidycensus (Walker 2019, Walker and Herman 2022, Moraga 2022) that allows for easy manipulation and quick visualizations. The tidycensus package also allows the user to browse through the tens of thousands of variables encompassed by the ACS. We added all of the data files to our project repository. All data and code for this project can be retrieved from the GitHub repository.

Table 1: Dataset Information

Data Sources	Variables Used	Time Period
Analyze Boston-City of Boston	Blue Bike Station Locations	Released in 2021
American Community Survey-U.S. Census Bureau	Percent of homeowners, Percent of residents carpooling, Percent of residents driving alone, Percent of residents using public transporation, Percent White, Percent Non-White	2016-2020 Census

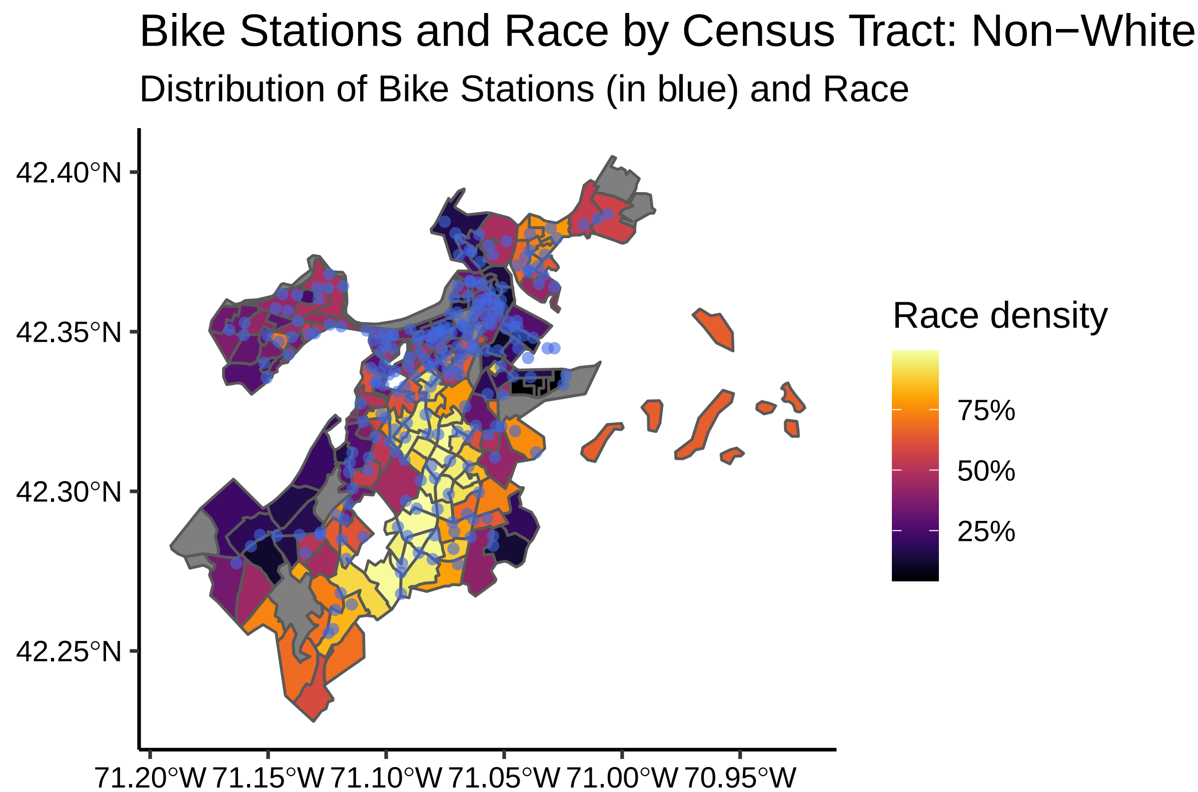
2.2 Data Wrangling

We began our analysis by importing data from both of our sources. For the Census data, we used the tidycensus package to select each of the variables in question individually. For the bike station data, we imported the shapefiles we had downloaded from the Blue Bikes Boston website. Next, we filtered all of the data on the county level from Suffolk County to the city of Boston. For the bike station data, this filtering was possible using the city name of Boston. For the tract-based Census data, this filtering was accomplished by selecting Boston-specific GEOID numbers. After importing the relevant ACS variables, we removed the margin of error column from the datasets, spread the data so the format would be conducive to the combining of datasets, and converted the raw numbers for several of the values of interest to percentages. We then converted the Census data to shapefiles, mapped a few of the Census variables against bike station locations for an initial look at the data, and combined the datasets using the "join" function to allow for combined regression analysis later on. The opening visualizations can be found in the "exploratory analysis" section below, along with our first impressions of those maps.

3 Exploratory Analysis

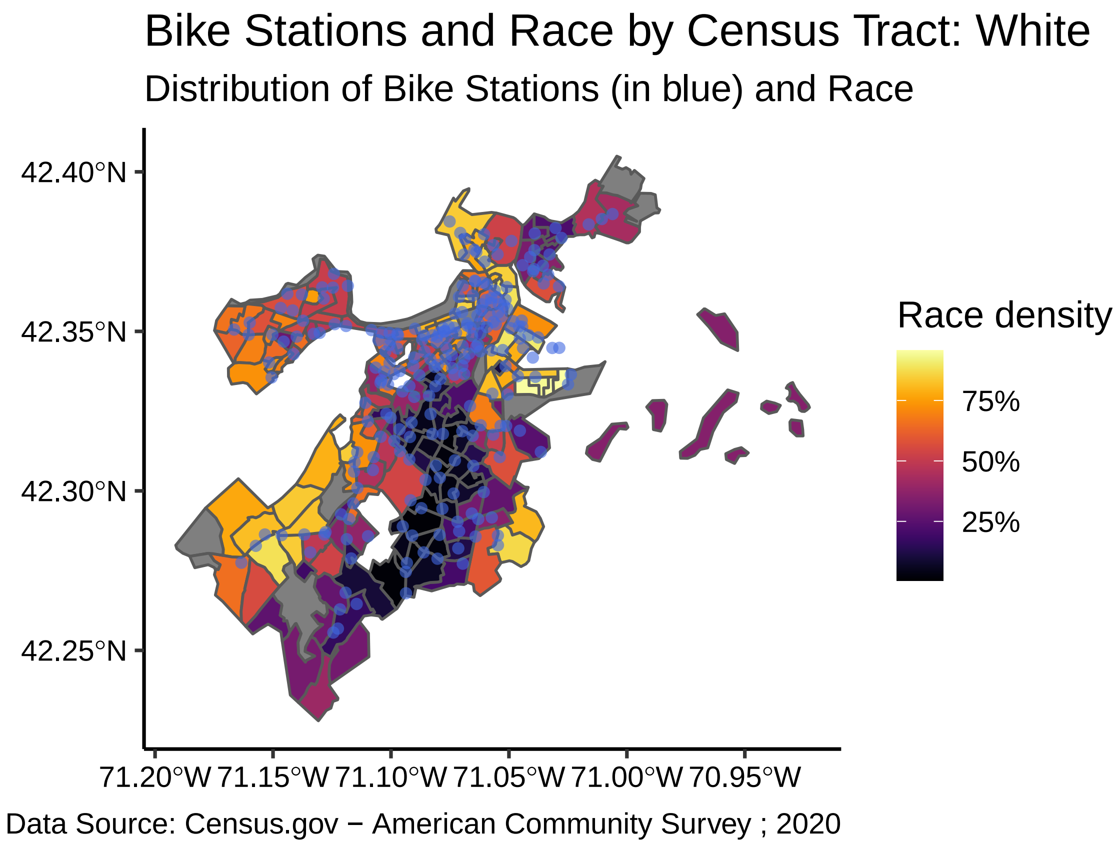
Following initial wrangling, we produced maps combining the locations of bike sharing stations and the depictions of certain key variables that relate to transit equity: income and race. We removed GEOIDs and neighborhoods that were listed as non-residential zones, as those areas would confound our analyses relating to homeownership as well as to commuting patterns. The results of this mapping—which informed our selection of additional variables that represent parts of the lives of urban families—can be found in the series of images below.

3.1 *Figure 1*

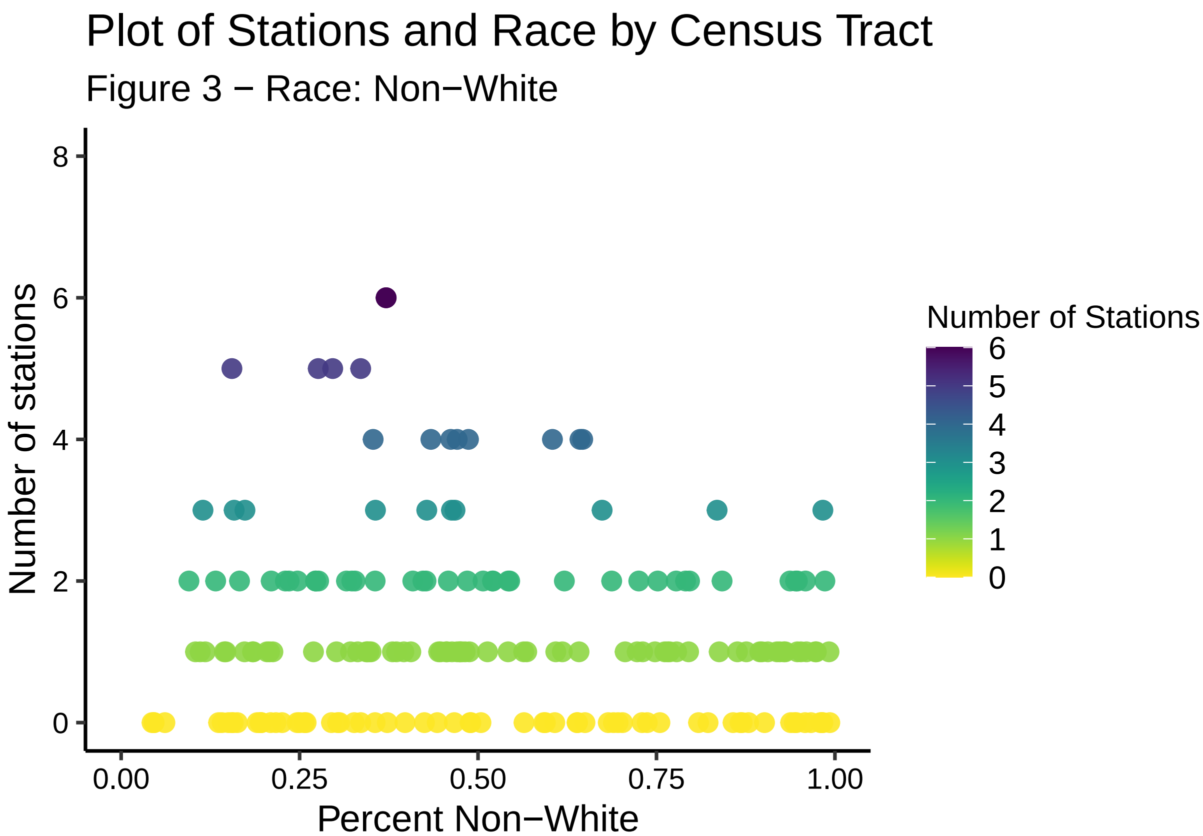


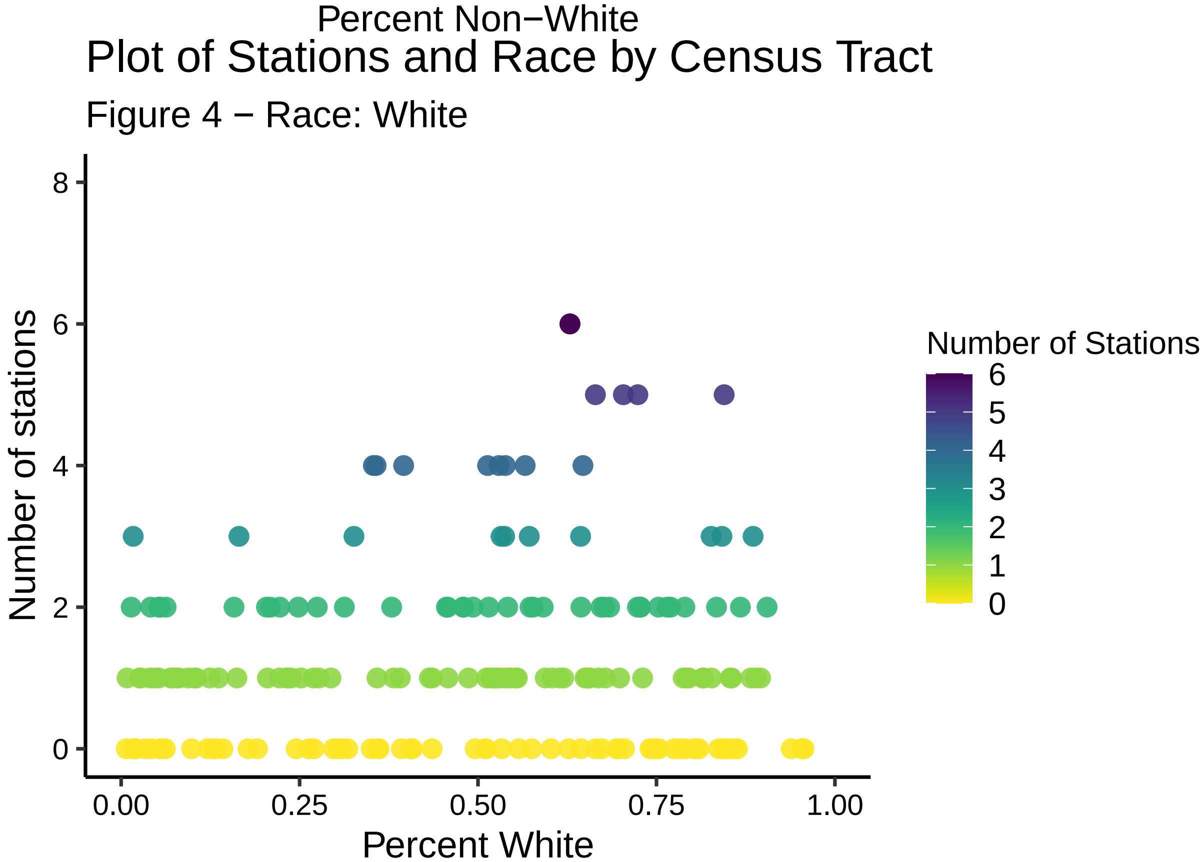
Data Source: Census.gov − American Community Survey ; 2020

3.2 *Figure 2*

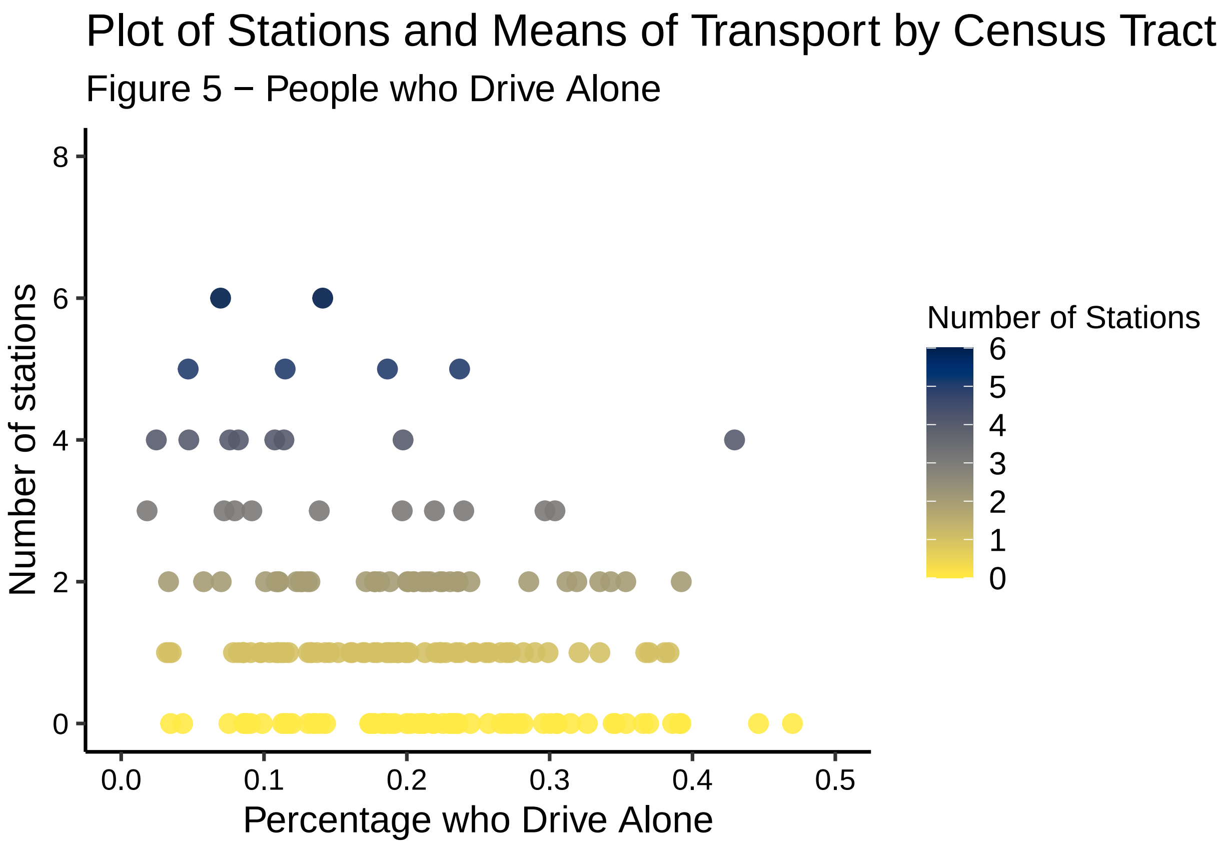

Bike Stations and Race by Census Tract: White
Distribution of Bike Stations (in blue) and Race
42.40°N
42.35°N
42.30°N
42.25°N
71.20°W
71.15°W
71.10°W
71.05°W
71.00°W
70.95°W
Race density
75%
50%
25%
Data Source: Census.gov − American Community Survey ; 2020

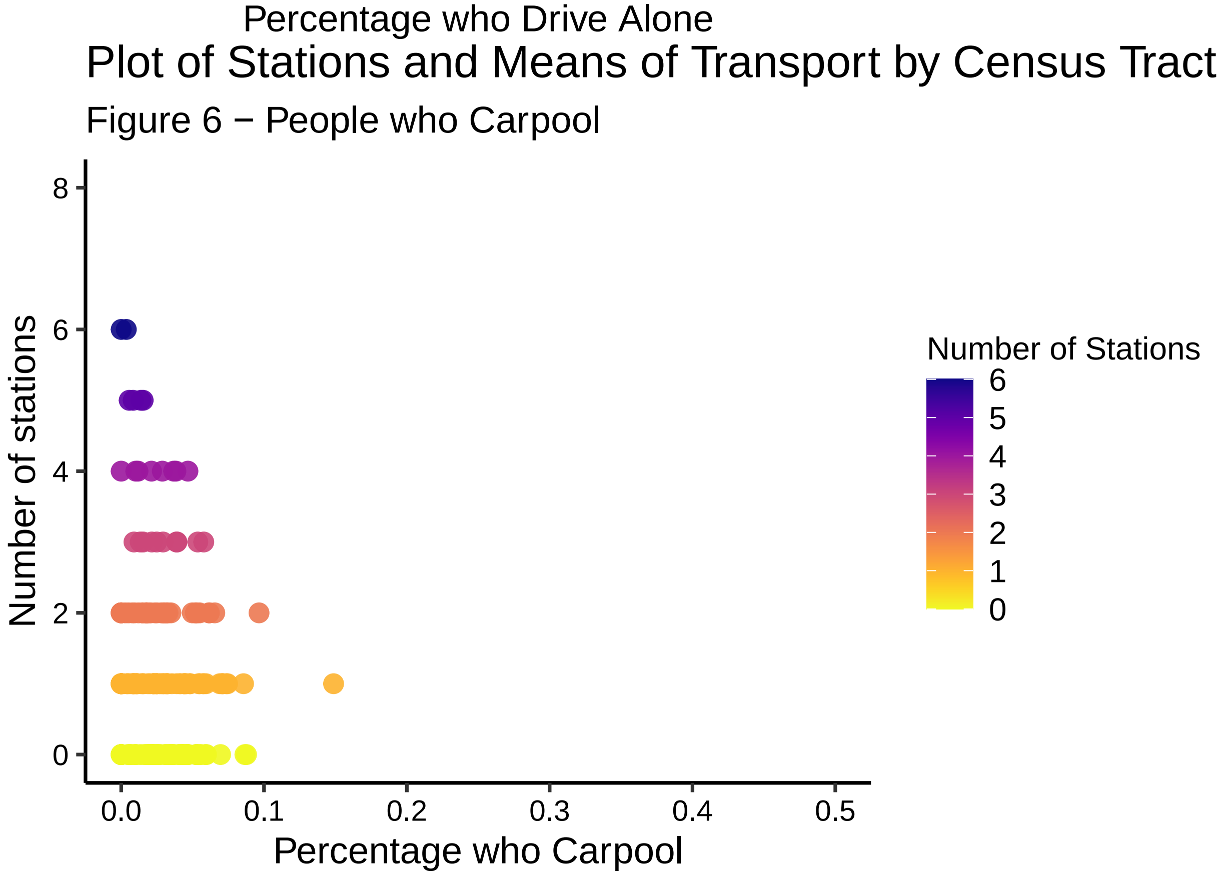
3.3 *Figures 3 and 4*



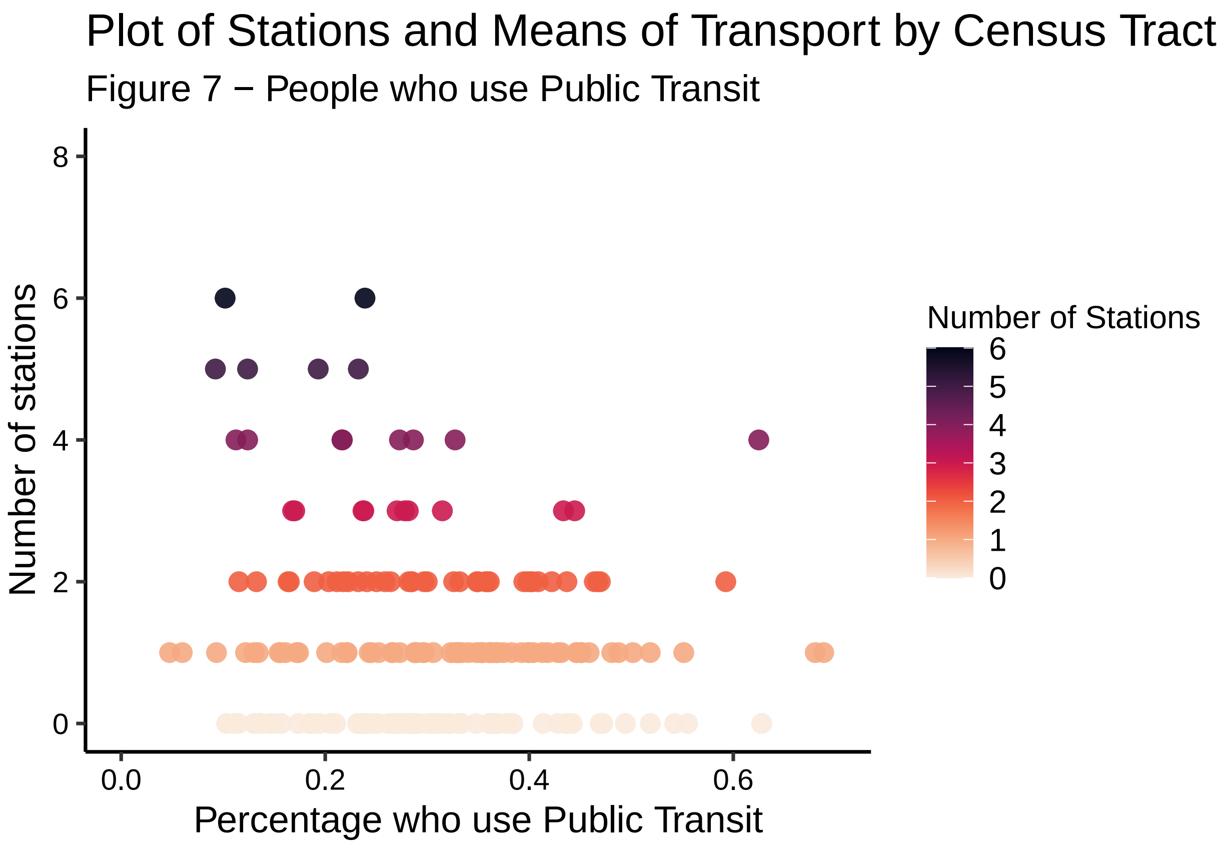


3.4 *Figures 5 and 6*





3.5 *Figure 7*



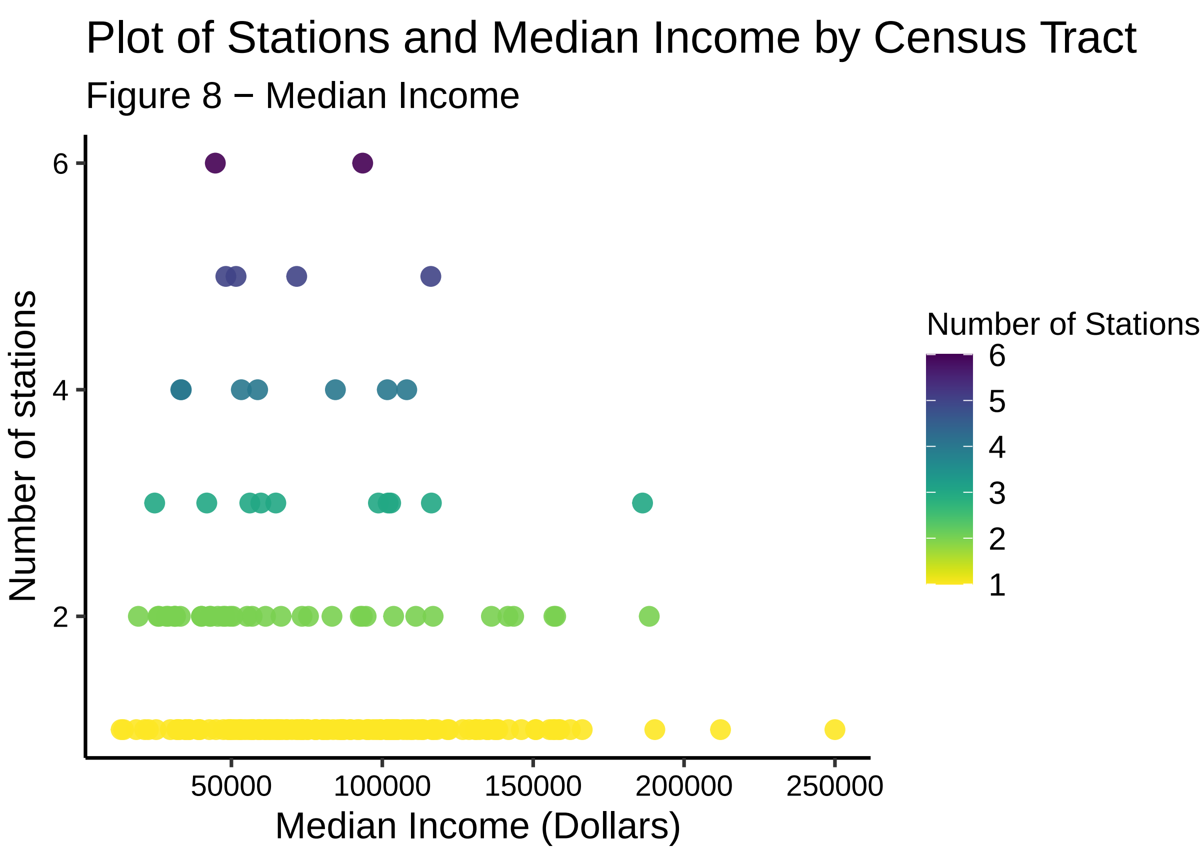
Plot of Stations and Means of Transport by Census Tract
Figure 7 − People who use Public Transit
Number of stations
8
6
4
2
0
0.0
0.2
0.4
0.6
Percentage who use Public Transit
Number of Stations
6
5
4
3
2
1
0

4 Analysis

4.1 Question: Are bike sharing stations in Boston equally available across population groups?

After mapping the most salient equity-related variables against the locations of bike stations in Boston, we parsed and prepared the other variables we had selected to be included in a General Linear Model (GLM). We put all the variables in one dataset and used an Akaike Information Criterion (AIC) call in R to isolate the variables that would best contribute to a robust combined regression. Then, we ran a regression and built a GLM with the strongest variables from the batch we had chosen. The variables with significance in our final regression were the percentages of people who use public transit to commute to work, and who commute to work by driving alone in a personal vehicle. It is worth noting, though, that the R-Squared values of our regression were low— 0.1129 for the Multiple R-Squared and 0.08865 for the Adjusted R-Squared—so the proportion of the variation in the locations of bike stations around Boston that is explained with our selected variables is low.

4.2 *Figure 8*



After mapping the most salient equity-related variables against the locations of bike stations in Boston, we parsed and prepared the other variables we had selected to be included in a General Linear Model (GLM). We put all the variables in one dataset and used an Akaike Information Criterion (AIC) call in R to isolate the variables that would best contribute to a robust combined regression. Then, we ran a regression and built a GLM with the strongest variables from the batch we had chosen. The variables with significance in our final regression were the percentages of people who use public transit to commute to work, and who commute to work by driving alone in a personal vehicle. It is worth noting, though, that the R-Squared values of our regression were low—0.1293 for the Multiple R-Squared and 0.1055 for the Adjusted R-Squared—so the proportion of the variation in the locations of bike stations around Boston that is explained with our selected variables is low.

5 Summary and Conclusions

5.1 Expected distribution inequity absent

In a typical example of a statistic representing whether bike sharing programs have been successful in reaching across user groups, one study found that in New York City, only 16.5% of people of color have access to Citi Bike, the city's bike sharing program (Pitt 2019). The literature also finds that the people who are likeliest to bike to work are young urban dwellers (US Census Bureau 2019), and "super-users" are more likely to be young, have household incomes below $75,000, and live and work near bike share stations (Winters et al. 2019). Our team expected to find similar results in Boston, another city that has experienced gentrification and the stratification of available amenities by socio-economic class and race. We did not find such results: both from the visual and the statistical perspective, demographic variables like income, race, and age do not have much to do with where bike share stations are in Boston. These findings could be because the barriers to participation that we imagined are not the ones that apply. A different obstacle could be perceptions of cycling safety, although riding a bike from a bike sharing program can be even safer than riding a personal bike (Walker 2019). For another idea, one author posits that lack of information, not access, is what makes bike share programs skew white and wealthy in many places (Schneider 2017)—in other words, many people just don't know enough about bike sharing programs to be comfortable taking advantage of them, despite their ready availability near where people live.

5.2 Alternate commute modes negatively correlate with Boston bike station locations

6 The variables that were significant in our regression of Boston bike station locations were the proportion of people who commute to work via public transit, and the portion who drive to work alone in a private car. For both variables, the coefficient in the regression was negative: -2.3935 at p = 0.00418 for public transit commuters, and -3.0081 at p = 0.01550 for those who drive to work in cars by themselves. Our model shows that the more people there are in a given census tract who commute to work either alone in personal vehicles or by public transit, the lower number of bike stations found in that census tract. This finding is not what we ancticipated: we posited that there would be a bundling of pro-environment indicators–more people using environmentally friendly commuting methods like public transit correlating with more bike stations available for people to use for errands, recreation, or commuting. Instead, the two significant variables in the regression point in opposite directions: fewer bike stations as more people commute to work alone in private cars (the bundling we expected), but also fewer bike stations as more people commute to work using public transit (not the bundling we expected). Perhaps the locations of Boston bike stations are irrespective of population characteristics altogether, and instead track other elements of the city, such as areas that are most popular with tourists. Also, the dataset we used for carpooling behavior was much smaller than we had hoped, raising the question of whether its sample size is sufficiently large to merit statistical confidence in the associated findings.The larger environmental context around bike sharing

6.1 Alternate commute modes negatively correlate with Boston bike station locations

The variables that were significant in our regression of Boston bike station locations were the proportion of people who commute to work via public transit, and the portion who drive to work alone in a private car. For both variables, the coefficient in the regression was negative: -2.3843 at p = 0.000332 for public transit commuters, and -2.1573 at p = 0.027758 for those who drive to work in cars by themselves (Table 2). Our model shows that the more people there are in a given census tract who commute to work either alone in personal vehicles or by public transit, the lower number of bike stations found in that census tract. This finding is not what we anticipated: we posited that there would be a bundling of pro-environment indicators–more people using environmentally friendly commuting methods like public transit correlating with more bike stations available for people to use for errands, recreation, or commuting. Instead, the two significant variables in the regression point in opposite directions: fewer bike stations as more people commute to work alone in private cars (the bundling we expected), but also fewer bike stations as more people commute to work using public transit (not the bundling we expected). Perhaps the locations of Boston bike stations are irrespective of population characteristics altogether, and instead track other elements of the city, such as areas that are most popular with tourists. Also, the dataset we used for carpooling behavior was much smaller than we had hoped, raising the question of whether its sample size is sufficiently large to merit statistical confidence in the associated findings.

Table 2: GLM Results

Variable	Estimate	P-Value	Significance
Percent White	0.2409	0.51150	
Percent Homeowner	-0.7069	0.21524	
Percent taking public transport	-2.3935	0.00418	**
Percent carpooling	-0.8204	0.85165	
Percent driving alone	-3.0081	0.01550	*

6.2 Alternate commute modes correlate with Boston bike station locations

The variables that were significant in our regression of where Boston bike stations are were those who commute to work via public transit, and those who drive to work alone in a private vehicle. For both variables, the coefficient in the regression was negative: -2.3935 at p = 0.00418 for public transit commuters, and -3.0081 at p = 0.01550 for those who drive to work in cars by themselves. Our regression tells us that those living nearer to bike stations are likelier to commute to work by public transit and less likely to commute to work alone in personal vehicles, but this relationship is a measure of correlation, not causation: the regression cannot prove a relationship with directionality between the variables. The larger environmental context around bike sharing programs asks whether these programs reduce the number of miles that urban residents drive, as people switch between driving and other modes of transport for their daily activities. Considering transportation is responsible for 27% of 2020 GHG emissions in US (EPA 2022), this question is an urgent one. Bike share programs have been found to reduce car use in some cities, but not all—and the effectiveness of bike share programs in protecting the environment is "dependent on whether [these programs replace] car use" (Fishman et al. 2014). With that said, bike share programs are most effective at curbing GHGs when they catalyze other changes, like increases in private bike use, improvements to bicycle infrastructure, and support of other policies that facilitate cycling (Schmidt 2018).

6.3 Limitations

The data we used is from the 5-year estimate from 2016-2019; therefore, this estimate primarily includes data from before the COVID-19 pandemic struck. Usage patterns have certainly changed since the pandemic—in some temporary, and some likely permanent, ways—and in fact there have been analyses performed on these changes. Because our analysis sought to investigate the baseline structure of the system, we worked with data from before the pandemic, but COVID-19 has transformed our society forever, so those changes are important to consider in future planning efforts.

Over the course our project, we encountered two primary analytical difficulties. The first was comparing results across regressions. A given independent variable might be significant in one regression and not significant in the other, even though the dependent variable of bike station locations was the same between both models. We could understand how the coefficient and significance might vary based on the other variables in the model, but it did not make sense to us that a variable that was statistically significant in one analysis could wholly drop from the statistically significant view in another model with the same outcome variable. The other analytical issue we faced was the types of variables available to us. The tidycensus package allows users to access and analyze the Census data quickly and easily, but it doesn't, of course, change what kinds of variables the Census includes. Even with our fairly wide distribution of variables, we still only explained about 10% of the variation in bike station locations in Boston. With more kinds of variables—say about land use types, business openings and closures, or consumer expenditures—we might capture more of the dynamics that help determine where these bike share stations get built.

6.4 Future Recommendations

As alluded to above, the first area we recommend future research target is incorporating other Census datasets and variables into this same type of analysis. Surveys of capital expenditures, retail trade, business formation and closures, and consumer expenditures could help illuminate the value flows in Boston that shape policymakers and developers perceptions and plans.

Secondly, it would be worth looking closely at how bike sharing impacts the health of user across neighborhoods in Boston. On a simplistic level, bike share programs improve users' health through more exercise (Clockston and Rojas-Rueda 2021), but that effect can be tempered by exposure to air pollution. Measuring the air pollution across different types of pollutants—especially PM 2.5, the most dangerous class—would complement the picture of user-level impacts of bike share programs.

Third and finally, the impact of bike share programs on overall city congestion would be an interesting direction for future analysis. One paper describes mixed effects of such programs on congestion in general, finding that “larger cities get better off while richer cities get worse off” and these programs “can reduce peak-hour congestion” (Wang and Zhou 2017). From both a quality-of-life and a GHG perspective, many different stakeholders would appreciate insight into the effects of the Boston Blue Bikes program on the congestion levels in the very packed city.

7 References

Bureau, U. C. (2019, May 14). *May 17 is National Bike to Work Day.* Census.Gov. https://www.census.gov/library/stories/2019/05/younger-workers-in-cities-more-likely-to-bike-to-work.html

Bureau, U. C. (2022, January 6). *About the American Community Survey.* Census.Gov. https://www.census.gov/programs-surveys/acs/about.html

Clockston, R. L. M., & Rojas-Rueda, D. (2021). Health impacts of bike-sharing systems in the U.S. *Environmental Research, 202*, 111709. https://doi.org/10.1016/j.envres.2021.111709

Fishman, E., Washington, S., & Narelle, H. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment, 31*, 13–20. https://doi.org/10.1016/j.trd.2014.05.013

Moraga, P. (2022). *Chapter 2 Spatial data and R packages for mapping | Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny.* Retrieved April 17, 2022, from https://www.paulamoraga.com/book-geospatial/sec-spatialdataandCRS.html

Motivate International Inc. (2022). *Bluebikes System Data.* Blue Bikes Boston. Retrieved April 17, 2022, from http://www.bluebikes.com/system-data

Plitt, A. (2019, July 10). Citi Bike fails to serve New Yorkers in poverty and communities of color: Report. Curbed NY. https://ny.curbed.com/2019/7/10/20689177/citi-bike-equity-expansion-report

Schmidt, C. (2018). Active Travel for All? The Surge in Public Bike-Sharing Programs. *Environmental Health Perspectives, 126*(8), 082001. https://doi.org/10.1289/EHP3754

Schneider, B. (2017, July 14). Bike Sharing Still Has a Stubborn Diversity Gap. *Bloomberg.Com.* https://www.bloomberg.com/news/articles/2017-07-14/explaining-the-diversity-and-equity-gap-in-bike-sharing

US EPA, O. (2015, December 29). *Sources of Greenhouse Gas Emissions [Overviews and Factsheets].* https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

Walker, A. (2019, December 16). *How bike share became the decade's biggest transit success story.* Curbed. https://archive.curbed.com/2019/12/16/20864145/bike-share-citi-bike-jump-uber

Walker, K. E. (2022). *Chapter 6 Mapping Census data with R | Analyzing US Census Data.* Retrieved April 17, 2022, from https://walker-data.com/census-r/mapping-census-data-with-r.html

Walker, K., & Herman, M. (2022). *Basic usage of tidycensus.* Retrieved April 17, 2022, from https://walker-data.com/tidycensus/articles/basic-usage.html

Wang, M., & Zhou, X. (2017). Bike-sharing systems and congestion: Evidence from US cities. *Journal of Transport Geography, 65*, 147–154. https://doi.org/10.1016/j.jtrangeo.2017.10.022

Winters, M., Hosford, K., & Javaheri, S. (2019). Who are the 'super-users' of public bike share? An analysis of public bike share members in Vancouver, BC. *Preventive Medicine Reports, 15*, 100946. https://doi.org/10.1016/j.pmedr.2019.100946