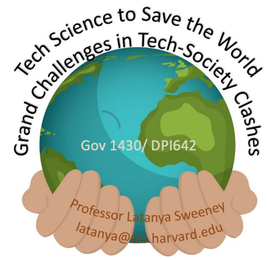
Are Amazon Hub Lockers equitably distributed?

Austin Nguyen, S.M. in Data Science, Harvard University IACS

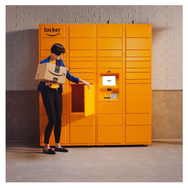
austinnguyen@g.harvard.edu



ABSTRACT

COVID-19 has accelerated the rise of e-commerce and home delivery, and automated parcel lockers are an avenue to improve delivery efficiency. Amazon Hub Lockers are secure, self-service kiosks that allow consumers to pick up their packages on their own time. While Amazon Hub lockers are strategically located, little research has been done on the accessibility and equity impacts. With data from the American Community Survey and Google Maps API, I model the count of Amazon Hub lockers and the presence of Amazon Hub lockers in Boston, MA following a regression framework. Given the wide consumer reach of Amazon and efforts to rebuff its public image from findings of racial bias in Same Day Delivery in 2016 [1], **my null hypothesis is that Amazon's placement of Hub lockers throughout the city of Boston do not exhibit racial bias.**

RESULTS SUMMARY



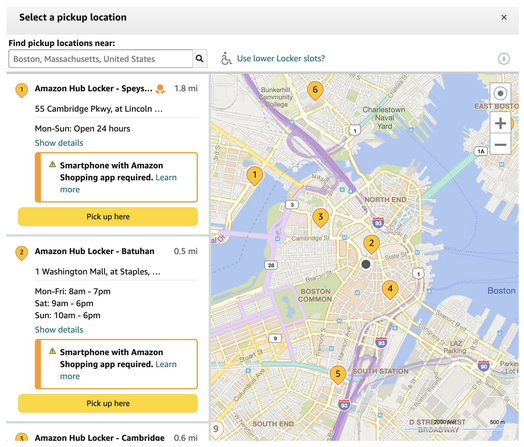
Race predictors, such as percentage of Black residents and percentage of white residents, are statistically significant (p-value < 0.01) and there is a stark differential in the coefficients (-10.06 vs. -6.79). Thus, I reject the null hypothesis that Amazon's distribution of Hub lockers in Boston do not exhibit racial bias. Given the environmental and economic advantage of Amazon Hub lockers, policymakers can encourage the expansion of last mile solutions to minimize market failures in underserved areas.

INTRODUCTION

Amazon Hub lockers are in more than 900 cities and towns across the U.S. While "strategically" located, are Amazon Hub lockers distributed equitably? Motivated by Amazon's recent controversies around historically black areas being less likely to have same-day delivery, this research builds on existing research on Amazon Hub locker distribution in Portland, OR [2] and Los Angeles, CA [3] to compare the distribution of Amazon Hub lockers in Boston against the racial breakdown of the city to understand whether any racial biases exist. I collect racial and household demographic data from the 2020 U.S. Census Bureau's American Community Survey (ACS) at the zip code level and the location of all Amazon Hub lockers using Google Maps API. I implement regression framework to understand the statistical significance of predictors on the count and presence of Amazon Hub lockers.

Data

The U.S. Census Bureau 2020 provides racial demographic information as well as median household income information in the Boston through the American Community Survey (ACS). The ACS is sent to approximately 295,000 addresses monthly or 3.5 million per year, thus making it the largest household survey administered by the U.S. Census Bureau. It gathers information on ancestry, citizenship, educational attainment, income, etc. The data is often used by public-sector, private-sector, and non-profit stakeholders to allocate funding and make decisions. From this data, I filter down zip codes to ones in Boston which result in 23 unique zip codes. Additionally, the Google Maps API provides a way to query for the name, address, latitude, and longitude of Amazon Hub lockers in Boston.



Methods

$$y = \beta_1 x_1 + \ldots + \beta_n x_n + \beta_0$$

I choose to formalize this research question in terms of a linear regression model and logistic regression model. In the linear regression framework, I model the count of Amazon Hub lockers in the Boston. In the logistic regression framework, I model the probability of the presence of Amazon Hub locker in a Boston zip code as framing the log-odds of having an Amazon Hub Locker. For the predictors, I examine the percentage of white residents, percentage of black residents, white majority zip code flag, median household income, and median household income, and total population in zip code as of 2020 census. This framework will allow us to estimate coefficients for each predictor which translates to estimate the effect of each predictor on the response.I standardize each predictor to ensure that predictors that have larger magnitude in scale do not disproportionate impact on the response

Results

Table 1:

	2 models	
	Linear Regression	Logistic Regression
pct_black	−10.06***	−9.5866*
	(2.227)	(5.657)
pct_white	−6.79***	−4.786
	(1.878)	(3.451)
white_majority	−0.9638***	−5.8201
	(1.015)	(6.186)
total_population_standardized	0.475**	2.9696
	(0.0223)	(1.774)
median_fam_income_standardized	0.0893	2.6587
	(0.345)	(2.308)
median_nonfam_income_standardized	−0.0397	−1.6515
	(0.319)	(1.984)
constant	8.1236***	6.784
	(1.673)	(5.604)
R-squared	0.715	0.5657
AIC	57.65	-
N	23	23

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

According a linear regression model, I observe a coefficient of -10.06 for the pct_black variable with a p-value < 0.01. This means that a one percent increase in Black individuals corresponds to a decrease of 10.06 in Amazon Hub lockers on average holding all other variables constant for a zip code in the Boston area. I observe a coefficient of -6.79 for the pct_white variable with a p-value < 0.01. Among the predictors that are not related to race, only total_population_standardized is statistically significant with a p-value of < 0.05. I also observe that the R-squared is 0.715, suggesting that the percent of variance explained in the number of Amazon Hub lockers in a zip code by this set of predictors is around 71.5%.

According to logistic regression framework in which the response is a binary outcome representing the presence of an Amazon Hub locker in 2022, I observe a coefficient of -9.58 for pct_black that is statistically significant with a p-value < 0.1 This can be interpreted that for every one percent increase in the percent of Black individuals in a zip code in Boston, I would expect a decrease of 9.58 in the log-odds of an Amazon Hub locker in that zip code holding all other variables constant. Among remaining predictors — white_majority, total population standardized, median family income standardized, median non-family income standardized, none are statistically significant. Lastly, the R-squared value is 0.565

Compared to the linear regression framework, I observe a weaker relationship between the predictor variables and the response. This may be a function collapsing the response variable of the number of Amazon Hub Lockers into a binary outcome, producing an response variable with less variability. Between the two models, pct_black provides to be a statistically significant predictor and both exhibit the same negative relationship to the response.

Discussion

I observe a statistically significant difference in the regression coefficients of the linear regression framing of this problem. The -10.06 coefficient for the pct_black and the -6.79 coefficient for pct_white both have a negative linear relationship with the response variable, but the magnitude of the negative linear relationship is different. Namely, I observe a larger decrease in the number of Amazon Hub lockers for pct_black compared to pct_white, with a difference of 3.27. **This finding gives evidence to reject the null hypothesis that the distribution of Amazon Hub lockers throughout the city of Boston does not exhibit racial bias.**

A limitation of this study is that the absence of interaction terms between race and household income. Given the history of racial segregation and red-lining in the United States and the structural barriers experienced by Black communities, I expect that income to vary with race.

Furthermore, the response variable has limitations as a metric of access. For example, the number of Amazon Hub lockers may all be concentrated in a specific area within a zip code. A more robust metric may be the distance to Amazon Hub lockers from each household or common commuting routes. Further work can explore other metrics of Amazon Hub locker accessibility.

NEXT STEPS

In stage 2, follow-up work could repeat the experiment but add features such as interaction terms such as income by race in zip code. Due to historic inequities, we would expect household income to depend on cofounder variables such as race and/or family type. Therefore, including household income by race directly as a predictor should minimize the impact of confounders on predicting whether a zip code has Amazon Hub lockers.

In stage 3, follow-up work would repeat the experiment for the greater Boston area (Cambridge, Somerville, Medford, etc) and explore other metrics of accessibility such as the distance between Amazon Hub lockers and every household in a zip code as a metric of accessibility.

REFERENCES

1. Ingold, David, and Spencer Soper. "Amazon Doesn't Consider the Race of Its Customers. Should It?" Bloomberg.com, Bloomberg, 21 Apr. 2016, www.bloomberg.com/graphics/2016-amazon-same-day/.
2. Keeling, Katherine; Schaefer, Jaclyn S.; and Figliozzi, Miguel, "Accessibility and Equity Analysis of Common Carrier Parcel Lockers at Transit Facilities in Portland, Oregon" (2020). Civil and Environmental Engineering Faculty Publications and Presentations. 574. https://pdxscholar.library.pdx.edu/cengin_fac/574.
3. Fang, Jiawen. "The Spatial Dynamics of Amazon Lockers in Los Angeles County". MetroTrans: Transportation Center. October 2019. https://www.metrans.org/assets/research/mf-5.4c_fang_final-report.pdf.

Harvard University Department of Government

HARVARD Kennedy School

Public Interest Tech Lab

DATA PRIVACY LAB