Exploring the Relationship Between Access and Usage of Public Transportation

Dad Science (cchrabas, lcohen16, gkern1, aeddy)

Hypothesis

High-quality public transportation has well-documented environmental and economic benefits and studies have suggested that improving access to it could make cities more inclusive and equitable. We sought to explore the impact of high-quality access to public transport in various communities. Specifically, our hypothesis is that there is a meaningful relationship between public transit access and usage. Further, we want to test whether this relationship varies by race.

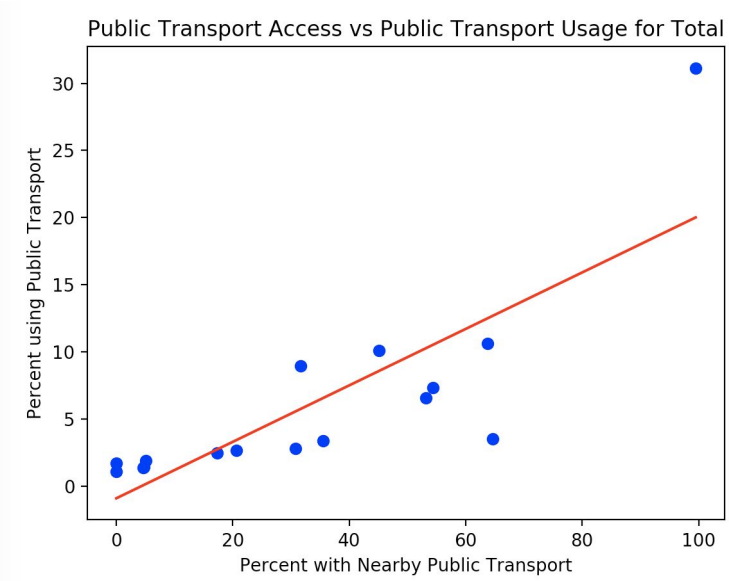
Data

We combined two datasets: "Transportation to Work" and "Walkable Distance to Public Transit". Both our datasets are from the California Department of Public Health (CDPH) whose data is from the U.S. Census Bureau, and were accessed through the public California Health and Human Services Open Data Portal. The cleaned dataset contains information by county on the percentage of people, stratified by race, living within a ½-mile of high quality public transportation (<15 min waiting time at peak commute hours), as well as the percentage of the county population who use public transit as a mode of transportation to work. Throughout our analysis, public transit access is defined as living within a ½-mile of high quality public transportation and public transit usage is limited to use as a mode of transportation to work.

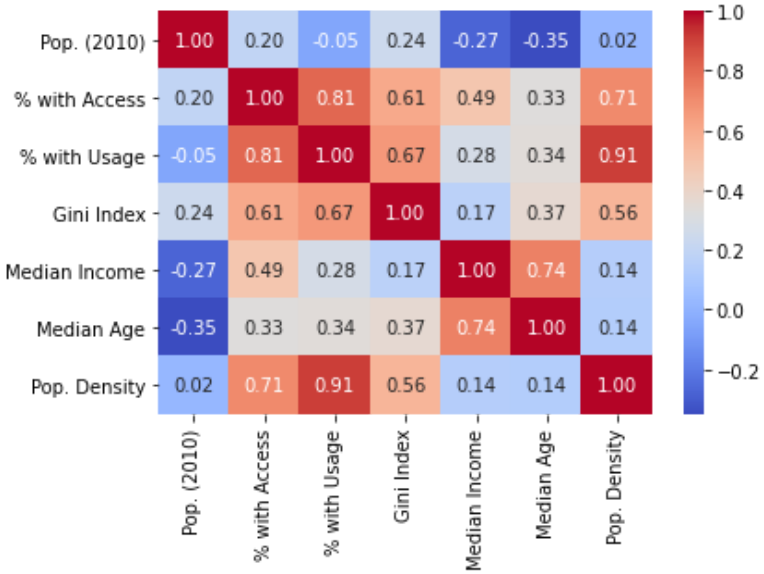
This sample is from 16 counties in California, which is comparably small to the 58 total counties in California. The transit usage dataset contained all 59 counties, but the transit access dataset contained only 19 counties, and 3 of those were missing correct dates. It is likely representative of California's residents and their transportation habits, but not of populations in the entire United States where public transportation systems vary. Sampling biases might influence our data in that people who are illiterate, ill, or otherwise incapable of filling out the U.S. Census (or unwilling or simply unaware of it) will not be represented in our data. Demographic data for the correlation matrix comes from the "Income Inequality" dataset from the California Health and Human Services Open Data Portal and the 2010 Demographic Profile for California counties from the U.S. Census.

Findings

Claim #1: People who have access to quality public transportation are more likely to use it.
We wanted to investigate the association between access to high quality public transit and public transit usage. To do this, we ran a regression analysis and found an R-squared of 0.81, a slope of 0.21, and a P value of 0.0001 (as seen below). For every additional percent of the county population with access to quality public transport, the average percent of the county population using public transit increases by approximately .21 percent.



We attempt to quantify how large an effect several confounding variables could have on this finding with the following Pearson's correlation matrix. We calculate the correlation coefficient between county population in 2010, percent accessible transportation (percent of the county population who live within 1/2 mile of high quality transit stop), percent usage transportation (percent of county residents aged 16 years and older who ride public transit to work), Gini index (measure of income inequality), county median income, county median age, and county population density.



Claim #2: There is a significant association between race and access to high quality public transit.

We wanted to investigate whether there is an association between race and access to public transit, which could be a confounding factor for Claim #1. To do this, we ran a chi-squared test of independence of variables in a contingency table of observed data, formatted as the total count of people of each race that have close access to public transit versus those who do not, to see if

the distribution of access to public transport was significantly different by race. We found that there was a significant difference (p=0.00), and we reject the null hypothesis that race and access to public transit are independent.
The average percentage of the population with and without access to quality public transit, divided by race, is illustrated in the chart below.

Race/Ethnicity	% With Access to Quality Public Transportation w/in ½ mile	% Without Access to Quality Public Transportation w/in ½ mile
AIAN	33.2%	66.8%
AfricanAm	39.2%	60.8%
Asian	33.1%	66.9%
Latino	40.0%	60.0%
Multiple	31.7%	68.3%
NHOPI	40.1%	59.9%
Other	38.3%	61.7%
White	27.9%	72.1%
Total	33.2%	66.8%

Claim #3: There is a significant association between race and usage of public transit to get to work.
We wanted to investigate whether there was an association between race and usage of public transit as a mode of transportation to get to work, which could be a confounding factor for Claim #1. To do this, we ran a chi-squared test of independence of variables in a contingency table of observed data, formatted as the total count of people of each race that use public transit versus those who do not, to see if the distribution of usage of public transport was significantly different by race. We found that there was a significant difference (p=0.00) and reject the null hypothesis that race and usage of public transit to get to work are independent.
The average percentage of the population that use and do not use public transportation to get to work, divided by race, is illustrated in the chart below.

Race/Ethnicity	% Using Public Transportation to Work	% Not Using Public Transportation to Work
AIAN	7.76%	92.24%
AfricanAm	8.91%	91.09%

Asian	6.13%	93.87%
Latino	8.15%	91.85%
Multiple	6.11%	93.89%
NHOPI	6.51%	93.49%
Other	6.04%	93.96%
White	4.73%	95.27%
Total	6.07%	93.93%

Claim #4: The association between public transit access and public transit usage to work does not vary by race.

We ran individual regression analyses on public transit access vs. public transit usage for each race and found the association does not vary significantly, because the 95% confidence intervals for each slope overlap with the 95% confidence interval of every other slope, as seen in the table below. We fail to reject the null hypothesis that the correlation between public transit access and public transit usage to work does not vary by race.

Regression Analysis on Association between Transit Access vs Transit Usage by Race

	Total	**White**	**Latino**	**AIAN**	**AfricanAm**	**NHOPI**
Slope	0.2102	0.2134	0.2272	0.2491	0.2144	0.2145
R Value	0.8113	0.8268	0.8015	0.6982	0.8132	0.6862
CI	0.13 - 0.29	0.14 - 0.29	0.14 - 0.32	0.12 - 0.38	0.13 - 0.30	0.08 - 0.35
P Value	0.0001	0.001	0.0002	0.0026	0.0001	0.0096