Exploring Access and Usage of Public Transit

Carlyn Chrabaszcz, Ailita Eddy, Lena Cohen, Garret Kern (Dad Science)

BROWN

Introduction

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

Original Datasets: "Transportation to Work" and "Walkable Distance to Public Transit" from the California Department of Public Health (CDPH)

Cleaned Dataset: Contains information by county on the percentage of people, stratified by race, with access to 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. 16 counties were included in the final dataset. 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.
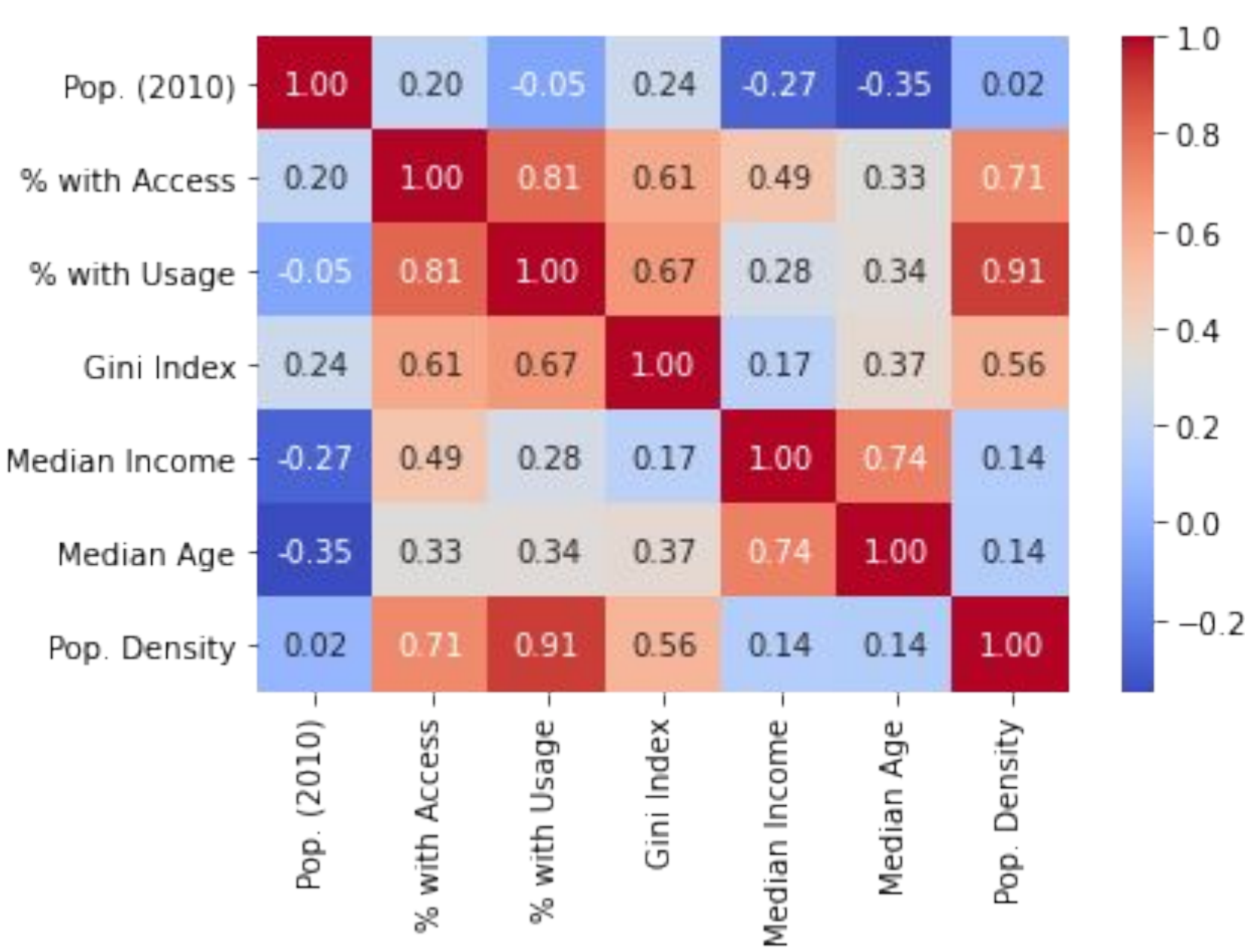
Data Specifications:
- Transit access is defined as living within walking distance of high quality public transportation
- Transit usage is limited to use as a mode of transportation to work.
- Sample is from 16 out of 58 counties in California

Sampling Biases:
- 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

#1: Access and Usage

We ran a regression analysis on public transport access vs. public transport usage to work and found an R^2 value of 0.81, a slope of 0.21, and a p-value of 0.0001 (Fig. 1). We conclude that there is a significant, positive relationship between living with a ½ mile of quality public transit and using public transit to get to work. We attempt to quantify the effect of several confounding variables of this relationship with the correlation matrix below (Figure 2).

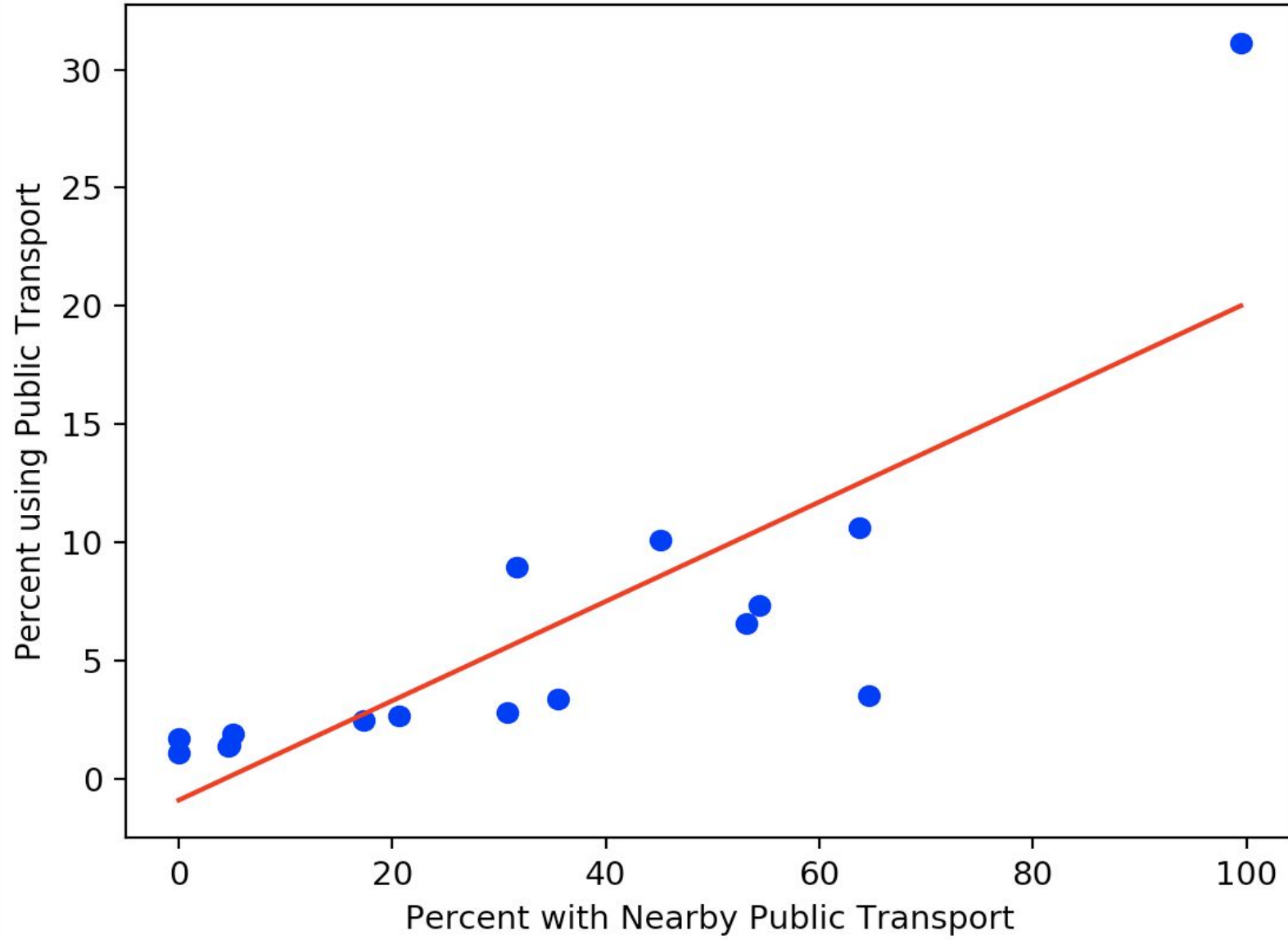Figure 1: Transport Access vs. Usage

Figure 2: Correlation Matrix

We display the Pearson's correlation coefficient between county population in 2010, county transit access percentage, county transit usage percentage, county Gini index (measure of income inequality), county median income, county median age, and county population density. Population density seems to be the confounding variable with the largest effect. The data is from the California Health and Human Services Open Data Portal and the U.S. Census 2010 Demographic Profile for California counties.

#2: Race and Access

We ran a chi-squared test 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 reject the null hypothesis that race and access to public transit are independent.

#3: Race and Usage

We ran a chi-squared test 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). We reject the null hypothesis that race and usage of public transit to get to work are independent.

Table 1: Distribution of Access and Usage

Race/Ethnicity	Access to Public Transport (%)	Use Public Transport to Work (%)
African American	39.2%	8.91%
American Indian/Alaska Native	33.2%	7.76%
Asian	33.1%	6.13%
Latino	40.0%	8.15%
Native Hawaiian/Other Pacific Islander	40.1%	6.51%
White	27.9%	4.73%

This chart displays the average across all 16 counties of the percentage of people who live within ½ mile of quality public transportation and the percentage of people who use public transportation to get to work, separated by race.

#4: Access and Usage By Race

We ran individual regression analyses on public transit access vs. public transit usage for each race and found the relationship does not vary significantly (Table 2). The confidence intervals for each slope overlap with the 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. Despite the association between race and access and race and usage, the linear relationship between access and usage is consistent across races.

Table 2: Regression Analysis on Access And Usage by Race

	AfricanAm	AIAN	Asian	Latino	NHOPI	White
Slope	0.214	0.249	0.219	0.227	0.215	0.213
R Value	0.813	0.698	0.788	0.802	0.686	0.827
CI	0.13 - 0.30	0.12 - 0.38	0.13 - 0.31	0.14 - 0.32	0.08 - 0.35	0.14 - 0.29
P Value	0.0001	0.0026	0.001	0.0002	0.0096	0.001

Discussion

We found a positive linear relationship between walkable access to quality public transportation and usage of public transit to get to work. The relationship between access and usage has implications for city planning. If a county expanding access to public transit leads to an increase in residents using public transit to get to work, studies suggest the environment will benefit and traffic congestion will be reduced. Access and usage varies across races, although the relationship between access and usage is not significantly different.

Future Work

Extensions of this project could examine the transportation usage and proximity data from all 58 California counties, or from all 50 states. Given that our data comes from 16 California counties, it likely does not represent trends surrounding public transit across the country. We would also like to continue exploring the relationships between confounding factors, such as wealth, income inequality, population density, health status, and age, and our two variables of interest, usage of public transit to get to work and walkable access to public transit. This could provide important insights into public transport usage trends across various demographic groups, which could help city planning decisions.