The Association of Public Infrastructure with Social Capital Measures

Aditya Koppikar
Beau Langdon

Leonard N. Stern School of Business
Glucksman Institute for Research in Securities Markets
Johannes Stroebel, PhD., Faculty Advisor
Theresa Kuchler, Director of Glucksman Institute

Abstract

Social Capital, defined as the strength of an individual's social network and community, has recently been studied by the Opportunity Insights research team for its relationship to upward economic mobility. Improving social capital for individuals potentially has ties to bringing people out of poverty as well as improving their health and overall quality of life. As a continuation of this research, in this paper, we would like to determine how various types of public infrastructure, such as libraries, museums, and farmers markets, are associated with multiple indicators of social capital, such as economic connectedness, clustering, and the volunteering rate of a community. Most notable in this research was that we found a statistically significant correlation between the average number of vehicles per household and many social capital indicators. Additionally, we found that many infrastructure variables displayed an initially positive/negative correlation followed by a reversal in the degree of the correlation with some of the social capital indicators, suggesting that there could be an efficient number of these institutions that maximize social capital for a community. Lastly, while we were able to find multiple correlations, further research would be required to establish causation between infrastructure and social capital.

Table of Contents

I. Introduction

Income inequality has become increasingly prevalent in the United States since 1980 (Horowitz et al., 2020). In this paper, we would like to determine which types of mediums and public infrastructure federal, state, and local governments can invest in to help foster an environment that promotes communication and friendship between people of various socioeconomic backgrounds to alleviate and combat income inequality. These institutions include libraries, fitness centers, farmers markets, and museums. Furthermore, we also regressed social capital measures against statistics on the percentage of people who walk to work and the average number of vehicles a family has. While this paper primarily establishes correlations between public infrastructure and social capital measures, and not causation, the results of this paper can help guide subsequent analyses and research of causal relationships.

We have regressed various public infrastructure quantities and attributes by zip code against social capital measures analyzed in the paper, "Social Capital I: Measurement and Associations with Economic Mobility," co-authored by Professor Johannes Stroebel and Professor Theresa Kuchler, among others. In this paper, "social capital," defined as the “strength of an individual's social network”, was analyzed (Chetty et al., 2020). This research stems from the Social Capital Atlas, a project led by Opportunity Insights at Harvard University. Opportunity Insights uses social network data to study social capital in the United States. In particular, data from 21 billion friendships is used to study social capital. The research above identifies social capital as a “potential determinant of various future outcomes, including education, employment, and health” (Chetty et al, 2020). There are multiple types of social capital, such as the “connectedness between people of low and high socioeconomic status groups” (economic connectedness), “the extent of cliques or groups forming in friendship networks” (social cohesion), and the “rates of volunteering” (civic engagement) (Chetty et al., 2020).

The rise in income inequality is something that the United States needs to manage urgently because lower-income groups will have to work increasingly harder to achieve similar results as their wealthier peers, leading to a more socially immobile society. As a result of income inequality, it is necessary to find mechanisms to support those who have not benefited as much economically in recent decades. There have been many programs and investments that the government and philanthropic institutions have funded to help reduce poverty in the United States. For example, the Earned Income Tax Credit (EITC) and Pell Grants offer tax credits and grants to low-income individuals to support them economically. However, while these solutions offer individuals a temporary boost in income, we must find solutions to help low-income individuals obtain sufficient opportunities on a more permanent basis to rise socioeconomically in society, a tenet of the American Dream.

Opportunity Insights has created a score for various social capital measures for each zip code in the U.S. The intended use of the Social Capital Atlas is "as a tool for researchers, policymakers, practitioners, and others to explore different forms of social capital in their communities, understand how these measures are associated with income mobility and other outcomes, and inform evidence-based policy solutions" (Opportunity Insights, 2022).

II. Literature Review

In addition to the Social Capital Paper, which was the motivation behind the research in our paper, research has been conducted around social capital for decades. Human capital theory can be traced to Jacob Mincer, who created a framework to determine the causes of inequality in personal income (Mincer, 1958). While the gross domestic product (GDP) has been seen as a standard economic indicator to determine the financial progress of a nation, in the paper, "Estimating the Mental Wealth of Nations," the authors propose the inclusion into a nation's wealth of several other factors that provide an indication of the health of a community, such as “employment conditions, stable housing, trust, and social connectedness cohesion” (Occhipinti et al., 2023). In the paper, "The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods," where a “large-scale early childhood intervention was conducted to analyze the educational attainment of over 2000 disadvantaged children in the US”, it was found that there are “significant spillover effects on children who lived near treated children” (List, Momeni, & Zenou, 2020). Therefore, spatial proximity matters to children's socioeconomic outcomes. In the paper "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," Miles Corak found that countries with greater income inequality tend to “have a greater fraction of economic advantage and disadvantage passed on between parents and their children” (Corak, 2013). This becomes increasingly important as income inequality measures in the US have steadily risen.

The share of American adults living in middle-class neighborhoods decreased from 61% in 1971 to 51% in 2019 (Horowitz et al.). Across the United States, geographic income inequality - the gap in income between richer and poorer places – “has risen more than 40% between 1980 and 1991” (US Department of Commerce, 2023). “Since 1981, the incomes of the top 5% of

earners have increased faster than the incomes of families of other groups" (Horowitz et al.). As a result, the share of US aggregate wealth has grown much faster for upper-income individuals than for middle-income or lower-income individuals. Additionally, neighborhoods are becoming increasingly segregated from an economic/income standpoint. In Nashville, the share of families living in middle-class neighborhoods "decreased by 15 percentage points between 1990 and 2020" (Kasakove & Gebeloff, 2022). Meanwhile, the proportion of people living in wealthy neighborhoods "jumped by 11 percentage points, while those who lived in poorer neighborhoods increased by four percentage points" (Kasakove & Gebeloff, 2022).

Many have turned to investments in education as a potential solution to helping to bring many in the US out of poverty. However, schools alone cannot help alleviate income inequality. In 1970, only about half of Americans ages 25 and older had a high school diploma, while as of 2019, 90% did (Hanauer, 2019). However, from 1979 to 2017, the average real annual wages of the top 1% of Americans rose 156% while the purchasing power of the average American's paycheck did not increase (Hanauer, 2019). Household income is the most predictive metric of a child's educational success (Hanauer, 2019). Even more so than the specific school a child attends, other aspects of their life affect their educational success, such as housing, healthcare services, and access to quality food.

One public infrastructure that has been analyzed in this paper is museums. The Museums Association and the Cultural Learning Alliance have conducted research that demonstrated that "learning through arts and culture improves attainment in all subjects" and "that students from low-income families who take part in art activities at school are three times more likely to get degrees" (Lewis, 2012). Furthermore, "students from low-income families who engage in the arts at school are more likely to volunteer and vote as young adults" (Lewis, 2012).

Another public infrastructure that has been analyzed in this paper is the public library. Public libraries provide society with access to information and are “uniquely positioned” to serve as a means of upward mobility for low-income groups, given that public libraries offer programming for education, workforce development, and employment opportunities (Winston, 2022). For example, libraries often have programs that teach children from an early age about employment opportunities, which can dramatically impact their lives. Furthermore, libraries for a community provide social opportunities for children and serve as significant mediums for building one's social network.

Fitness centers have been analyzed in this paper due to recent trends that may contradict earlier hypotheses on the association between fitness centers and income inequality. While fitness centers and gyms were often seen as places for those of all walks of life to interact at the bench press, treadmill, or group fitness class, the rise in luxury fitness businesses, such as “Equinox, Barry's Bootcamp, and Orange Theory”, has been a significant factor in the expansion in the fitness industry. These luxury fitness centers comprise about “40% of the existing fitness market” (McNeill, 2022). In response to this, there has also been a rapid growth in “budget gym chains, such as Planet Fitness” (McNeill, 2022). Therefore, similar to what was discussed earlier regarding a decrease in the middle class, there has been a decrease in “mid-price” fitness centers (McNeill, 2022). This trend may prevent interactions between people of different socioeconomic backgrounds from interacting with one another. Therefore, simply adding a fitness center to a community may not improve social capital measures; it may often depend on the type of fitness center added to the community.

Another public institution that can help improve the state of the community's mental and physical health is the farmers market. As of 2019, “40 million Americans lived in food

insecurity", without consistent access to healthy food (USC Bedrosian Center, 2019). People in these socioeconomic backgrounds can only choose from what is locally available. Therefore, farmers markets are increasingly important to neighborhoods because they sell fresh produce and local delicacies directly to consumers and allow people to be better connected to their communities and to the farmers who are bringing them their food. Many farmers markets accept government food stamps and SNAP benefits (USC Bedrosian Center, 2019), making high-quality food affordable for low-income individuals. Therefore, farmers markets serve not only as a medium for residents to access high-quality, nutritious food but also for interacting with other community residents from various socioeconomic backgrounds.

In this research, we analyzed two separate metrics that represent their means of transportation: the percentage of a zip code that walks to work and the average number of cars a household owns. These metrics were chosen because the nation has become increasingly divided regarding how people commute to work. It appears that "larger, denser metros" with "more educated, affluent people" are more likely to walk to work or use public transit (Florida, 2019). On the other hand, a higher concentration of working-class jobs is "positively associated with driving alone to work and negatively associated with using transit or walking to work" (Florida, 2019). However, having access to a car also provides opportunities for low-income families to access the aforementioned public infrastructure (libraries, museums, farmers markets, and fitness centers). Therefore, while walking to work may be a privilege afforded to the rich, it may also be necessary to track how vital having access to a car is for low-income families to be able to access mediums that will allow them to achieve upward income mobility.

Therefore, in this paper, the public infrastructure we have assessed against social capital measures are libraries per capita, museums per capita, fitness centers per capita, farmers markets

per capita, the percentage of a zip code that walks to work, and the average number of cars owned by a household.

III. Data Methodology

As discussed earlier, we have analyzed the social capital measures defined in the paper, “Social Capital I: Measurement and Associations with Economic Mobility”. These categories were developed on a “dataset of 72.2 million users of Facebook aged between 25-44 years, involving 21 billion friendships, in the United States”. These online Facebook friendships serve as proxies for real-life friendships and relationships.

The social capital measures employed in this paper as the response variables include Economic Connectedness, Neighborhood Economic Connectedness, Neighborhood Bias, Clustering, Civic Organizations, Support Ratio, and the Volunteering Rate. The variables above all contain data at a zip code level. The definitions provided by the Social Capital Website are listed in **Figure 1**.

Figure 1: Definitions of Social Capital measures that are analyzed in this paper

(These definitions are directly from the Social Capital Paper)

Variable Names (dataset name)	Description
Economic Connectedness (ec_zip)	Baseline definition of economic connectedness: two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the ZIP code
Neighborhood Economic Connectedness (nbhd_ec_zip)	Economic connectedness calculated using only within-neighborhood friends.
Neighborhood Bias (nbhd_bias_zip)	nbhd_ec_zip divided by nbhd_exposure zip, all subtracted from one. This represents the tendency for low-SES people to befriend high-SES people conditional on exposure to high-SES people
Clustering (clustering_zip)	The average fraction of an individual's friend pairs who are also friends with each other.
Civic Organizations (civic_organizations_zip)	The number of Facebook Pages predicted to be "Public Good" pages based on page title, category, and other page characteristics, per 1,000 users in the ZIP code.
Support Ratio (support_ratio_zip)	The proportion of within-ZIP code friendships where the pair of friends share a third mutual friend within the same ZIP code.
Volunteering Rate (volunteering_rate_zip)	The percentage of Facebook users who are members of a group which is predicted to be about 'volunteering' or 'activism' based on group title and other group characteristics.

Source: Social Capital I: measurement and associations with economic mobility

More information on how these variables were calculated can be found in the codebook for Publicly Available Data on Social Capital (Chetty et al., 2020)

The following explanatory variables are used in this paper. This data has been sourced from PolicyMap, a mapping platform that contains data on demographics, real estate, health, schools, housing, employment, and public investments.

Figure 2: Definitions of Public Infrastructure/Independent variables that are analyzed in this paper

Explanatory Variable (dataset name)	Description
Libraries per Capita (lib_per_capita)	The number of libraries in a zip code per 100,000 people
Fitness Centers (fitness_centers)	The number of fitness centers in a zip code per 100,000 people
Farmers Markets (farmers_markets)	The number of farmers markets in a zip code per 100,000 people
Museums (museums)	The number of museums in a zip code per 100,000 people
Percentage of People who Walk to Work (percentage_walk)	The percentage of people in a zip code who walk to work
Number of Vehicles (num_vehicles)	The average number of vehicles per household in a zip code

Source: Datasets from PolicyMap (Works Cited Sources 2-7)

All of the explanatory and response variables have been "Winsorized," meaning that the top and bottom 5% of the datasets have been assigned the score of the capped percentiles (5% for the lowest data points and 95% for the highest data points). We have done this to control for significant outliers within the datasets.

Control Variables:

In **Figure 3** have controlled for the following variables within our binned scatter plots and regressions: mean income, college education, log of the population, and median age.

Figure 3: Definitions of Control Variables

Control Variable (dataset name)	Description	Data Source
Mean Income (mean_income)	The mean income of working individuals in this zip code	Census Bureau Data
College Education (bachelors_degree_or_more)	The proportion of the population that has a college degree or higher	Census Bureau Data
Population (log_population)	The natural log (base e) of the population of the zip code	Social Capital Dataset
Median Age (median_age)	The median age of people in this zip code	Census Bureau Data
Population Density (pop_density)	Estimated Number of People per Square Mile	Census Bureau Data

Source: PolicyMap, Social Capital Dataset (Works Cited Sources 8-11)

The population and the median age variables datasets were also "Winsorized." The mean income, college education variables, and population density variables were not "Winsorized" since they did not contain significant outliers.

For each analysis, we conducted a regression of each of the seven social capital measures described above against each of the six explanatory/independent variables. We also included the control variables, mean income, college education, log population, and median age, for each regression between one of the social capital measures against one of the independent variables.

The regressions are conducted using the Binsreg function in R, which showcases the relationship between two variables based on partitioning/binning the independent variable of interest into ten deciles. This provides a pictorial representation of the relationship between the social capital measures and the infrastructure variables.

Furthermore, we conducted multivariate regressions of the social capital measures against the independent variables (including the control variables). For the multivariate regression, we replaced the log population variable with the population density variable.

The results for the univariate and multivariate regressions have been shown in the next section.

IV. FINDINGS

A. Regression Analysis Results

Figure 4: Summary of the Univariate Regressions between the Social Capital Measures and the Public Infrastructure variables

	Economic Connectedness	Neighborhood Economic Connectedness	Neighborhood Bias	Clustering	Support Ratio	Volunteering Rate	Civic Organizations
Libraries per Capita	**3.97E-04*** **	**4.86E-04*** **	**2.77E-04**	**-2.51E-04*** **	**-9.23E-04*** **	**-2.37E-04*** **	**-1.10E-05**
Fitness Centers per Capita	**-5.69E-04*** **	**-1.22E-03*** **	**8.04E-04*** **	**-1.65E-05**	**-1.74E-04**	**-2.54E-04*** **	**1.68E-04*** **
Farmers Markets per Capita	**3.65E-04**	**6.68E-04*** **	**-1.62E-03***	**-3.96E-04*** **	**-5.23E-04**	**-1.35E-04**	**-4.29E-05**
Museums per Capita	**7.21E-05**	**-3.50E-04** **	**5.89E-04*** **	**-1.19E-04*** **	**-3.80E-04*** **	**-7.17E-05*** **	**1.05E-04*** **
Percentage of People Who Walk to Work	**1.34E-03*** **	**-3.81E-03*** **	**7.03E-03*** **	**2.82E-04*** **	**1.60E-03*** **	**7.88E-04*** **	**1.30E-03*** **
Average Number of Vehicles	**8.51E-02*** **	**1.26E-01*** **	**-1.21E-01*** **	**-9.50E-04*** **	**3.57E-02*** **	**1.71E-02*** **	**-3.58E-03*** **

*****, p<0.01; **, p<0.05; *, p<0.10.**

The table above summarizes the univariate correlations between the independent variables and the various social capital indicators. For example, the coefficient on the libraries per capita variable when the economic connectedness variable was regressed on it was found to be 3.97 $x10^{-4}$. It was statistically significant to the 1% level. However, the coefficient for the fitness centers per capita variable in a regression with the economic connectedness variable was -5.69 x 10^{-4}. In these univariate regressions, each of these social capital measures was regressed against the independent variables one at a time, with only the control variables included in the

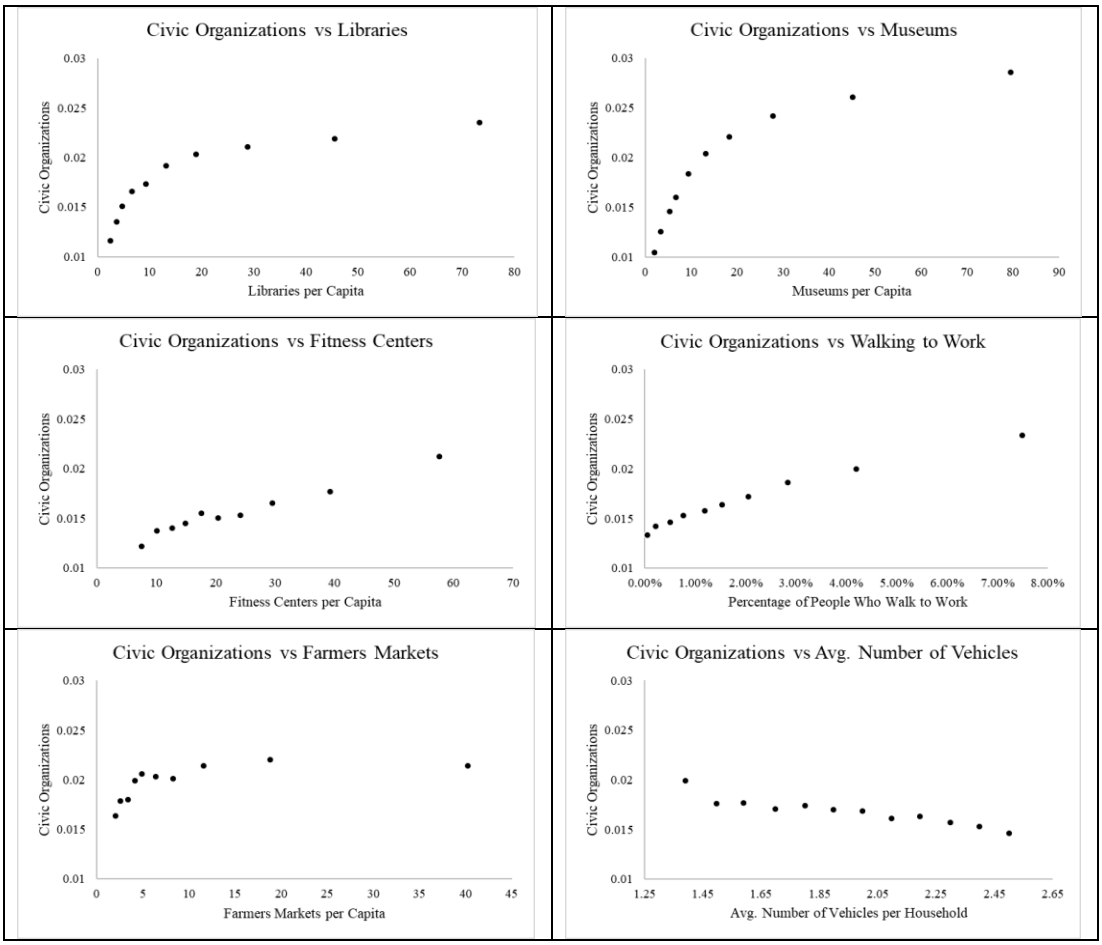
univariate regressions to control for variation in the age, income, education, and population of a zip code.

Additionally, this table indicates the level of statistical significance for each of the 42 associations. As shown above, the percentage of people who walk to work and the average number of vehicles owned per household have a statistically significant association with all the social capital measures. The neighborhood economic connectedness social capital measure is the most likely, among the seven social capital measures, to be able to be affected by the inclusion or exclusion of institutions in a neighborhood (e.g., libraries, museums), given that all of the independent variables displayed a statistically significant association with the social capital variable. However, this would require further research, and the research conducted in this paper primarily establishes correlations and not causal relationships between the social capital measures and the public infrastructure.

B. Binned Scatter Plot Analysis Results

Plots are organized by their social capital variables, however an alternate view is provided in the Appendix, where plots are grouped together by the independent variables.

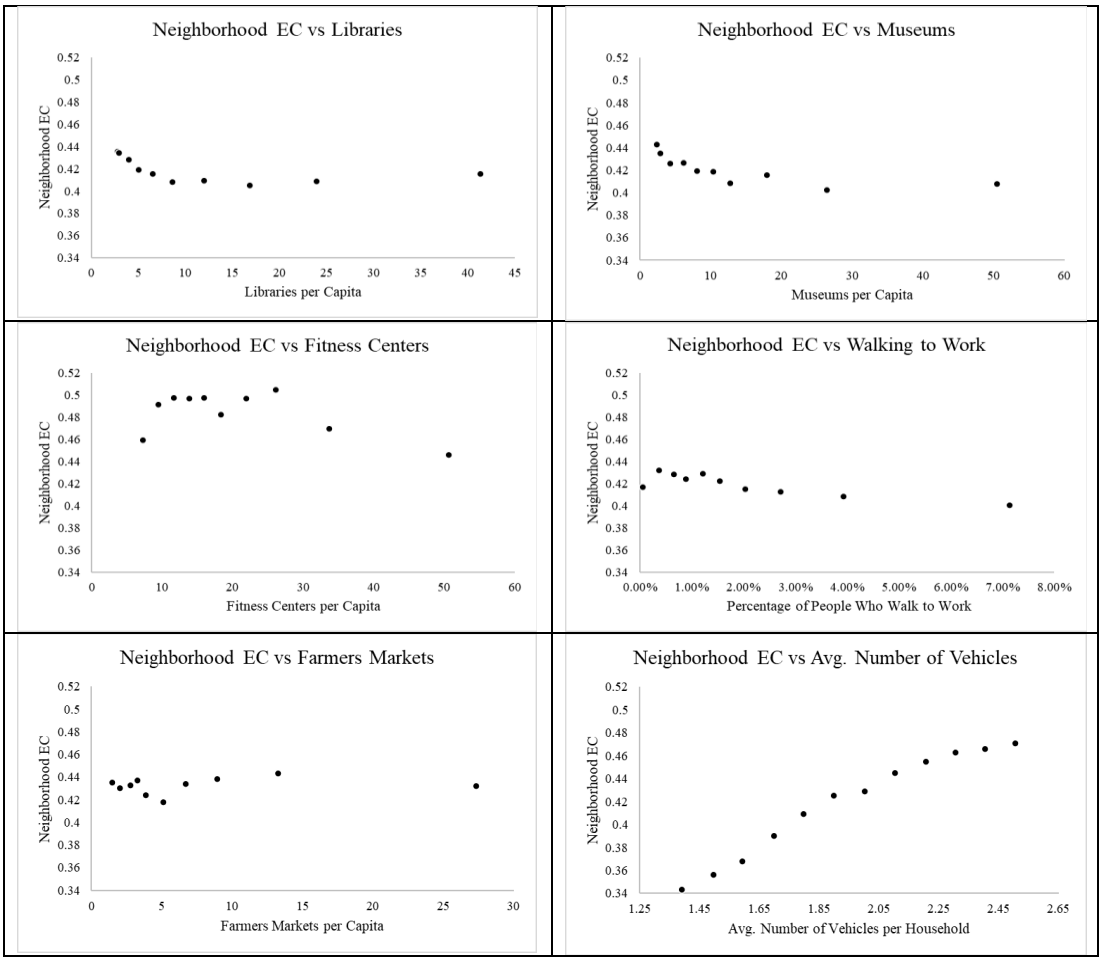
Figure 5: Scatter plots of the Civic Organization social capital measure against the public infrastructure variables



Overall, the libraries per capita, museums, fitness centers, and the percentage of the population that "walks to work" strongly and positively correlate with the number of civic organizations in the zip code. The number of farmers markets per capita has a slight positive association with the number of civic organizations in a zip code. However, there is a peak at about 20 farmers markets per zip code, after which there is a negative association. The average number of vehicles per household has a negative association with the number of civic organizations in a zip code. Another point to note is that we see relatively smooth positive

correlations for both libraries and museums, which should be expected because both libraries and museums are considered civic organizations.

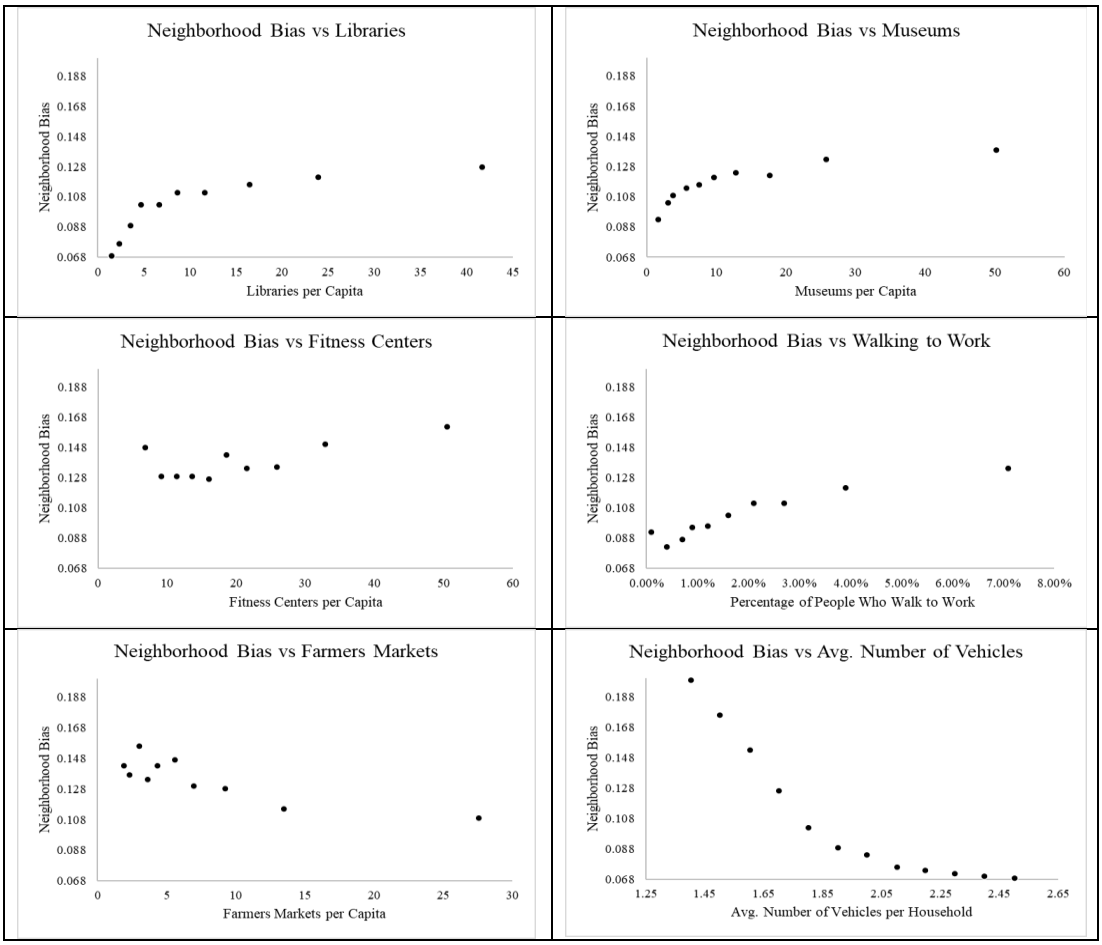
Figure 6: Binned scatterplots of the neighborhood economic connectedness social capital measure against the public infrastructure variables



The average number of vehicles per household was the only variable among the six potential indicators positively associated with a zip code's neighborhood economic connectedness. The number of libraries per capita, museums per capita, fitness centers per capita, and the percentage of people who walk to work in a zip code all have a negative association with neighborhood economic connectedness. However, instead of showing a consistent decrease throughout the ten deciles, there are peaks and troughs, after which the data

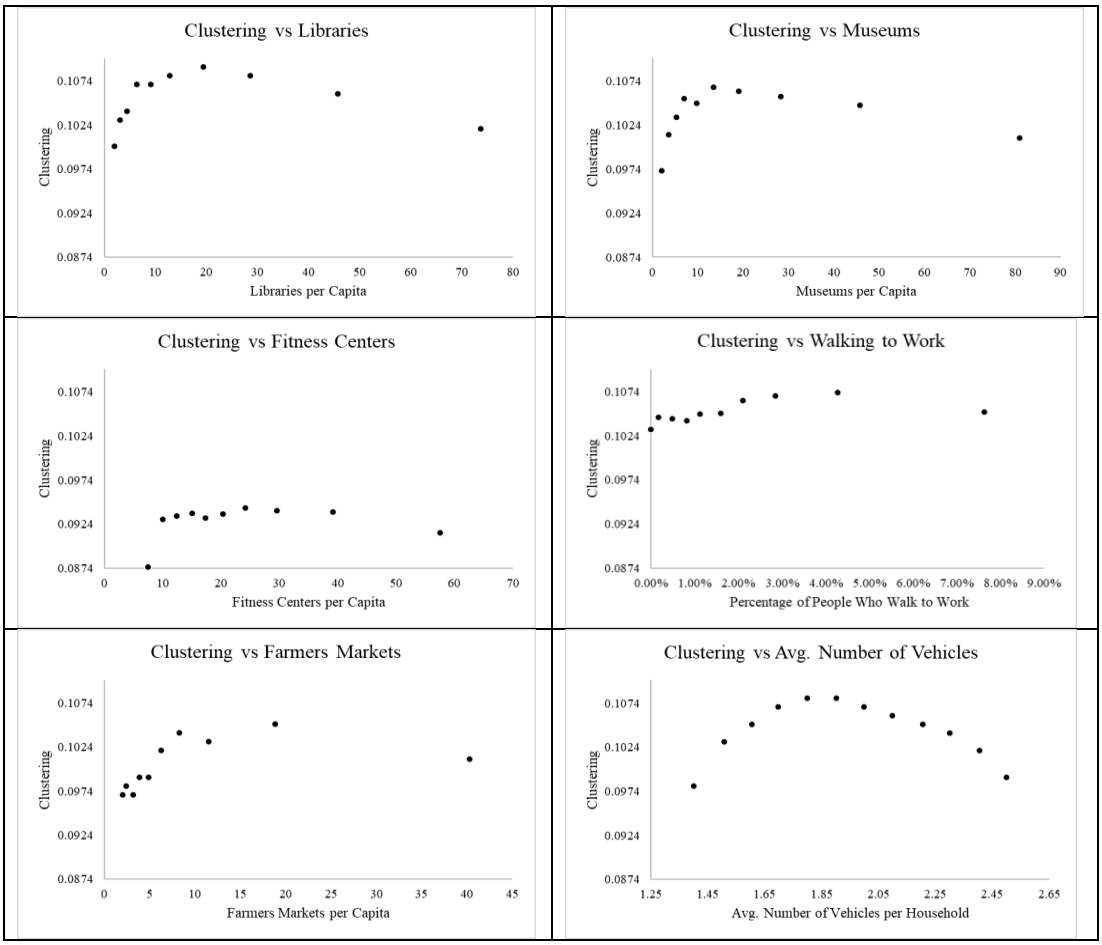
goes in the opposite direction. For example, for libraries per capita, the neighborhood economic connectedness decreases until about an average of under 20 per 100,000 residents, after which an increase in libraries is associated with an increase in neighborhood economic connectedness. For museums per capita, the neighborhood economic connectedness decreases until about a little over 25 libraries per 100000 residents, after which an increase in museums is associated with an increase in the neighborhood economic connectedness. There is a positive association between neighborhood economic connectedness and fitness centers until about 30 fitness centers per 100,000 people, after which the association becomes negative. The percentage of people who walk to work peaks at about the second decile, after which the association is negative. The association between the neighborhood's economic connectedness and farmers markets fluctuates between positive and negative throughout the various deciles in the binned scatterplot.

Figure 7: Binned scatterplots of the neighborhood bias social capital measure against the public infrastructure variables



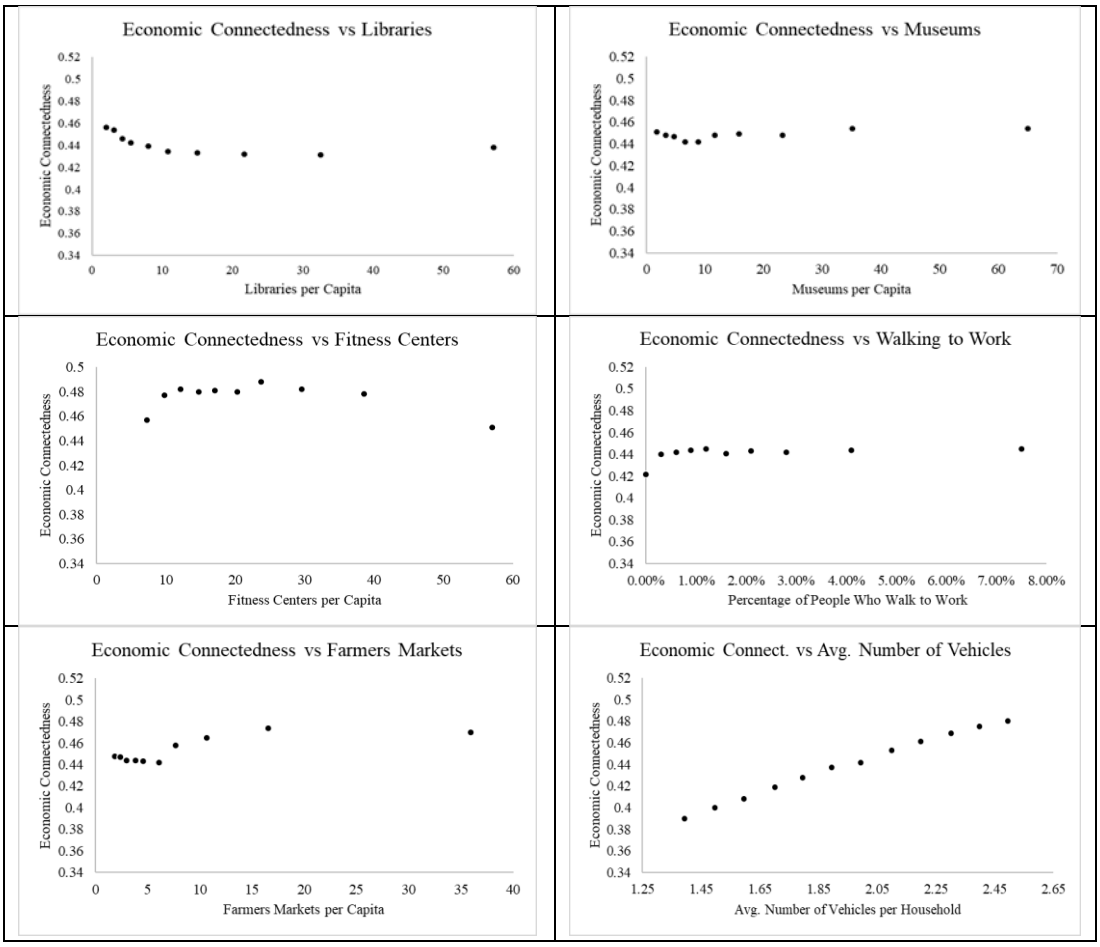
The associations between the independent variables and the neighborhood bias are apparent from the pictures above. Libraries per capita, museums per capita, and the percentage of people who walk to work strongly correlate with the neighborhood bias in a zip code. Fitness centers per capita have a generally positive association with neighborhood bias. However, the association fluctuates between a positive and a negative association for the middle deciles. The farmers market per capita and the average number of vehicle variables have negative associations with that neighborhood bias.

Figure 8: Binned scatterplots of the clustering social capital measure against the public infrastructure variables



The associations between the variables and the clustering indicator initially increase before decreasing rapidly after the peak level. The level of clustering in a zip code peaks at about 20 libraries per 100,000 residents. The level of clustering in a zip code peaks at about 13 museums per 100,000 residents. The level of clustering in a zip code peaks at about 25 fitness centers per 100,000 residents. The level of clustering in a zip code peaks at a rate of about 4% of the population walking to work. The level of clustering in a zip code peaks at about 19 farmers markets per 100,000 residents. The level of clustering in a zip code peaks at an average rate of about 1.875 cars/household in a zip code.

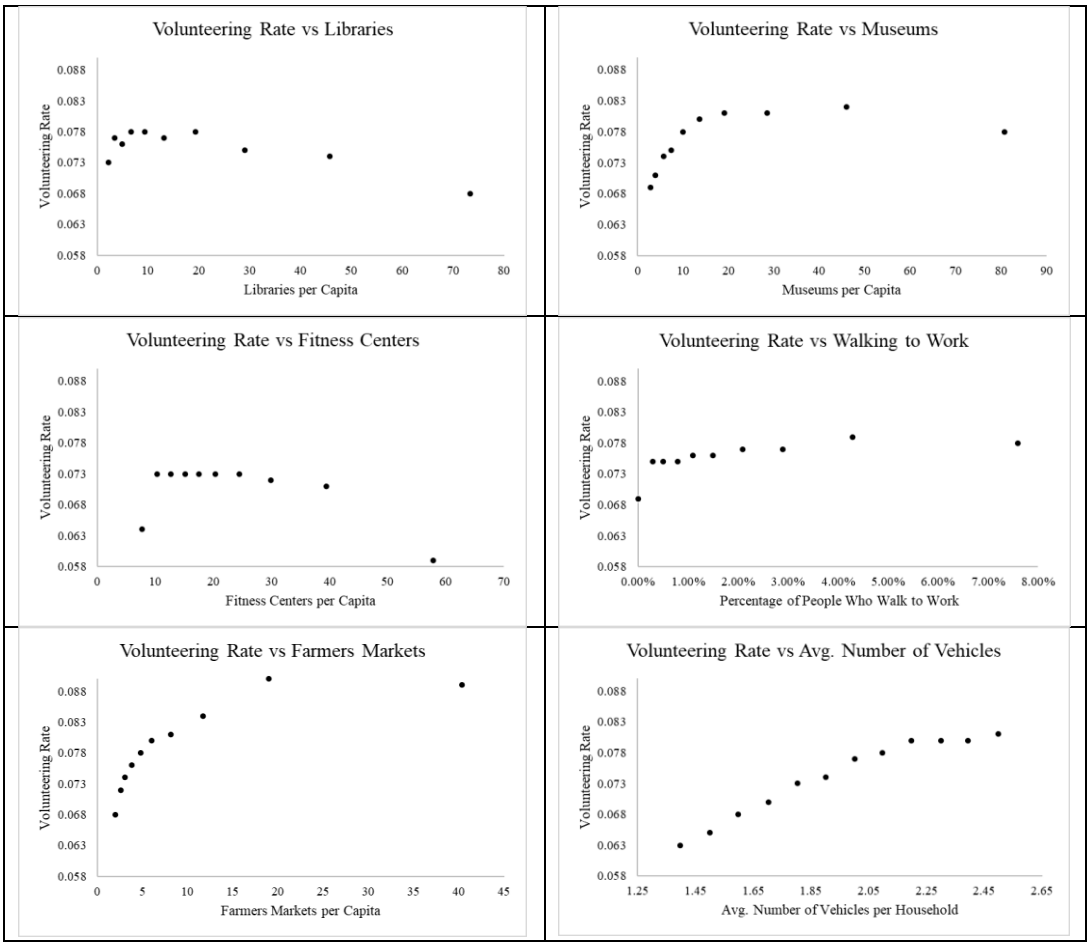
Figure 9: Binned scatterplots of the economic connectedness social capital measure against the public infrastructure variables



The only variable with a clear positive association throughout the various deciles is the average number of vehicles per household, where an increase in the number of cars is consistently associated with an increase in the economic connectedness of a zip code. The number of libraries per capita in a zip code is negatively associated with the economic connectedness of a zip code, although the economic connectedness increases after the ninth decile. The association between museums per capita and economic connectedness initially decreases before increasing. The association between fitness centers per capita and economic

connectedness initially increases to a peak economic connectedness level at around 25 fitness centers per 100,000 residents before decreasing. The economic connectedness level jumps dramatically between the first and the second decile for the variable relating to the percentage of people who walk to work, after which there is a small positive association between the two variables. The association between the farmers markets per capita and the economic connectedness of a zip code decreases until a trough of around 7 farmers markets per 100,000 residents, after which the association dramatically increases.

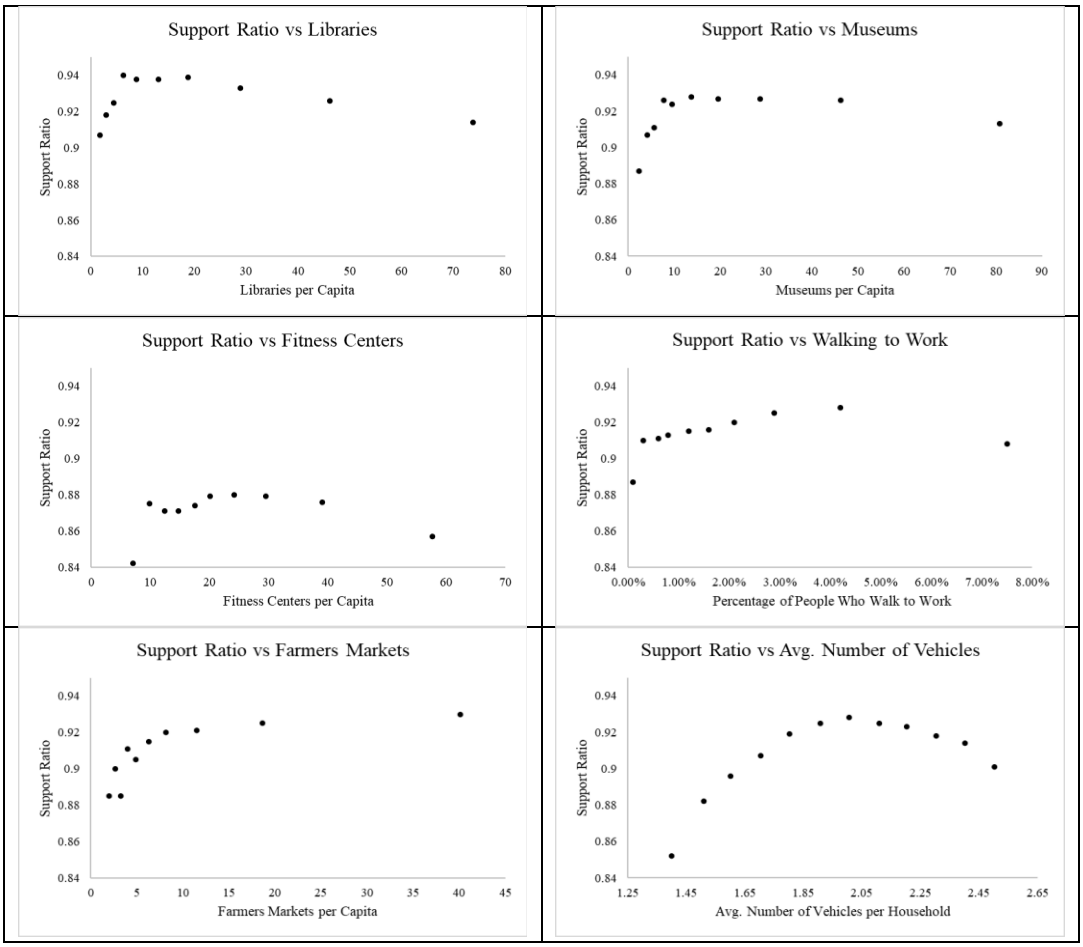
Figure 10: Binned scatterplots of the volunteering rate social capital measure against the public infrastructure variables



The association between libraries per capita and the volunteering rate of the zip code is initially positive until a peak volunteering rate at around 19 libraries per 100,000 residents, after which the association becomes negative. The association between museums per capita and the volunteering rate of the zip code is initially positive until a peak volunteering rate at around 45 museums per 100,000 residents, after which the association becomes negative. The association between fitness centers per capita and the volunteering rate of the zip code is initially positive until a peak volunteering rate at around 21 fitness centers per 100,000 residents, after which the association becomes negative. The association between the percentage of the zip code that walks to work and the volunteering rate of the zip code is initially positive until a peak volunteering

rate at around 4% of the population that walks to work, after which the association becomes negative. The association between farmers markets per capita and the volunteering rate of the zip code is initially positive until a peak volunteering rate at around 19 farmers markets per 100,000 residents, after which the association becomes negative. The average number of vehicles per household is positively associated with the volunteering rate of the zip code throughout the ten deciles.

Figure 11: Binned scatterplots of the support ratio social capital measure against the public infrastructure variables



The association between libraries per capita and the support ratio is initially positive until a peak support ratio at around 8 libraries per 100,000 residents, after which the association becomes negative. The association between museums per capita and the support ratio is initially positive until a peak support ratio at around 13 libraries per 100,000 residents, after which the association becomes negative. The association between fitness centers per capita and the support ratio is initially positive until a peak support ratio at around 24 fitness centers per 100,000 residents, after which the association becomes negative. The association between the percentage of people who walk per capita and the support ratio is initially positive until a peak support ratio

at around 4.25% of the population that walks to work, after which the association becomes negative. The association between farmers markets per capita and the support ratio is generally positive throughout the deciles of the plot, though fluctuating in the middle deciles. The association between the average number of vehicles per capita and the support ratio is initially positive until a peak support ratio at around 2 cars per household, after which the association becomes negative.

C. Multivariate Analysis

Figure 11: Multivariate Analysis of the Social Capital measures against the Public Infrastructure variables

	Economic Connectedness	Neighborhood EC	Neighborhood Bias	Clustering	Support Ratio	Volunteering Rate	Civic Organizations
Libraries per Capita	-1.00E-03	-5.82E-04	3.05E-03*	-2.98E-05	-1.60E-04	-1.10E-04	1.40E-04
Fitness Centers per Capita	-1.85E-04	-1.73E-04	-6.40E-04	1.15E-05	-3.70E-04	-4.30E-04***	9.63E-05*
Farmers Markets per Capita	-2.46E-03*	3.59E-03	-7.00E-03**	7.74E-05	-3.09E-03***	6.30E-04	-8.58E-06
Museums per Capita	-3.69E-04	2.01E-04	-2.20E-04	-1.44E-05	1.26E-04	-3.39E-05	2.120E-04***
Percentage Walk to Work	7.48E-03***	-2.39E-03	2.04E-03	4.11E-04	2.24E-03	2.68E-03***	1.51E-03***
Average Number of Vehicles	1.57E-01***	2.15E-01***	-1.20E-01***	-4.07E-03	-2.24E-02	1.67E-02**	-3.56E-03
Mean Income	-1.08E-07	1.31E-06***	-4.55E-07	-1.95E-08	3.65E-07***	-2.59E-07***	3.42E-08
Bachelor's Degree or More	4.12E-01***	5.55E-01***	-8.92E-02	-5.64E-03	-2.21E-01***	1.20E-01***	-4.27E-03
Population Density	4.37E-06**	1.16E-05***	3.26E-06	-2.51E-06***	-2.34E-05***	-1.41E-06*	-1.31E-06***
Median Age	1.11E-03*	-6.29E-04	1.33E-04	3.82E-05	1.24E-03**	5.09E-04**	4.30E-04***

*****, p<0.01; **, p<0.05; *, p<0.10.**

The above table shows the multivariate analysis results, where the social capital variables were regressed on all of the explanatory variables discussed in this paper and the control variables. One change made was a regression on population density as a variable rather than a

log population variable. Only a few variables are statistically significant when we include all of the explanatory and control variables in the regression. The most statistically significant explanatory variable is the average number of vehicles owned, followed by the percentage of people who walk to work. On the other hand, it appears that libraries per capita or museums per capita only exhibit one strong correlation each with the social capital measures analyzed in this paper. Fitness centers per capita and farmers markets per capita variables are moderately statistically significant with the social capital variables.

None of the independent variables exhibited statistically significant correlations with the clustering variable. Regarding the average number of vehicles per household, it appears that increasing this metric for a zip code is associated with both positive outcomes and negative outcomes. For example, an increase in the average number of vehicles in a zip code is associated with greater economic connectedness, greater neighborhood economic connectedness, less neighborhood bias, a higher volunteering rate, and less clustering. However, an increase in the average number of vehicles per household in a zip code is also associated with a lower number of civic organizations.

V. Discussion

Both the regression results and the various binned scatterplots indicate how the different types of public infrastructure can be associated with social capital measures.

For the univariate regressions, the regression results show that neighborhood economic connectedness is the social capital measure that has the potential to be impacted the most (among the social capital variables analyzed in this paper) by changes in the public infrastructure within a zip code. Furthermore, farmers markets per capita has the least statistically significant associations with the social capital measures presented in this paper, given that the associations

between farmers markets and four of the seven social capital measures were not statistically significant at a 10% level. However, analyzing the binned scatterplots provides more granular detail of the associations between the various independent variables and the social capital measures. There is a significant association for many variables until a certain point, after which the association reverses. For example, for libraries per capita, the economic connectedness decreases as the number of libraries increases until a level of about 32 libraries per 100,000 residents, after which increasing the number of libraries per capita is associated with an increase in the economic connectedness of a zip code. For another example in the research, an increase in fitness centers is associated with a decrease in neighborhood bias until a level of about 15 fitness centers per 100,000 residents, after which an increase in the number of fitness centers per 100,000 residents is associated with an increase in neighborhood bias. Therefore, it seems that for these potential government interventions, there is a peak/trough level at which social capital is maximized. This makes sense since, for example, you would want sufficient libraries in a zip code to provide resources to an entire region but not so many as to indirectly make some libraries more accessible for wealthier communities and neighborhoods than others.

Furthermore, it appears that, even within the same public infrastructure medium, increasing the amount of the variable per capita can be associated with positive outcomes in one social capital measure while simultaneously being associated with negative outcomes in another. For example, it was shown in the binned scatterplots that increasing the number of museums in a zip code is associated with a higher degree of economic connectedness. However, an increase in museums is also associated with higher neighborhood bias and more clustering.

After conducting the multivariate regression analysis, it became apparent that the explanatory variable that is the most statistically significant with the social capital measures is

the average number of vehicles per household. Even after controlling for many factors, an increase in the average number of vehicles per household is associated with an increase in positive outcomes, such as economic connectedness and the neighborhood economic connectedness, the volunteering rate, and a decrease in negative outcomes, such as neighborhood bias and clustering. However, an increase in the average number of vehicles per household is also associated with a decrease in the number of civic organizations, which would be detrimental to improving the connectedness of a group of people. Owning a car, or more cars, may be important to improving social capital because an additional car enables a person or a family to access more opportunities, whether it be job opportunities, educational opportunities, or social opportunities. These are all ways people can improve their socioeconomic status in a community. It may be possible that by owning a car, a person will have more opportunities to interact with higher-income individuals, which may be a reason why an increase in cars is associated with an increase in social capital measures.

With the data as it stands, we cannot analyze the results and determine causal relationships between the explanatory variables and the social capital variables. This research is primarily an observational study, where we observed and analyzed the relationships between different variables without any interventions. To determine whether there are causal relationships in the study, we would have to conduct experimental research or an experimental study. This would be tough to conduct across the entire United States and would involve recalculating the social capital measures after a specific number of years and then determining the change in the explanatory variables compared to the current experiment and to see how that change has affected the social capital measures. This would then serve as a proxy for an experimental study and could provide some indication of causal relationships.

While we included various control variables, there may be different types of control variables we can include in future research. Instead of just including a median age variable, we could include indicator variables for the various generations (e.g., Silent Generation, Baby Boomers, Gen X, Millennials, Gen Z) to see if variation in ages across a zip code is more or less likely to be associated with various social capital measures, such as the volunteering rate or the clustering rate. Similarly, it would have been good to dissect the income variable into multiple income levels corresponding to low, middle, and high income to analyze how the addition of various public infrastructure mediums in a region differs based on the composition of income groups in a region.

If we were to continue this research, we could include more variables that highlight further aspects of the lifestyle of the US population. We could include transportation variables corresponding to the average travel time to work and distance to the nearest public transit station to better understand how access to transportation modes can be associated with social capital measures. We could also analyze the number of jobs per capita within a zip code to determine how the job opportunities of an area helps bridge people of various socioeconomic statuses. We could also analyze educational variables, including the number of private schools, community colleges, and 4-year colleges in a zip code, to determine how adding various academic institutions in a zip code can help improve social capital measures. A better understanding of the association of educational institutions and social capital measures is important because public schools tend to receive more government funding and will be associated with higher enrollment from lower-income individuals than private schools. A health variable to analyze further would be the number of physician offices per capita to determine how better access to medical care and

doctors, which is typically the case for high-income individuals, can help improve a community's social capital measures.

Therefore, for future research, it would be crucial to further analyze not just the analyzed public infrastructure but additional ones as well to determine what it is about libraries, museums, etc., that increasing the number per capita may be associated with an increase in a social capital measure to a certain point, after which any increase could be associated with negative outcomes for the various same social capital measure. Future research would involve better understanding the independent variables analyzed in this paper. For example, what types of programs are offered in the libraries, what kinds of events are held at the museum, how expensive are the groceries offered at the farmers markets in a given area, etc. With answers to these questions, we would have more information that could help governments promote a more socially mobile and equitable society.

Works Cited

1) Horowitz, J.M. , Igielnik, R., Kochhar, R (2020, January 9). Views of Economic Inequality. Pew Research Center. www.pewresearch.org/social-trends/2020/01/09/views-of-economicinequality

2) Institute of Museum and Library Services (2018). Museums. PolicyMap. https://www.policymap.com/. https://www.imls.gov/research-evaluation/data-collection/museum-data-files

3) Institute of Museum and Library Services (2021). Public Libraries. PolicyMap. https://www.policymap.com/. https://www.imls.gov/research-evaluation/data-collection/public-libraries-survey/explore-pls-data/pls-data

4) United States Department of Agriculture (2022). Farmers Markets. PolicyMap. https://www.policymap.com/. https://www-ams-usda-gov.proxy.library.nyu.edu/local-food-directories/farmersmarkets

5) Dicennial Census and American Community Survey (2017-2021, last updated January 2023). Estimated percent of workers who commuted to work by walking in 2018-2022. https://www.policymap.com/.https://data-census-gov.proxy.library.nyu.edu/cedsci/

6) Dicennial Census and American Community Survey (2017-2021, last updated January 2023). Estimated Average Number of vehicles per household in 2018 - 2022. https://www.policymap.com/.https://data-census-gov.proxy.library.nyu.edu/cedsci/

7) Census County Business Patterns (2003-2019). Count and Percent of certain health-related establishments. https://www.policymap.com/. http://www.census.gov.proxy.library.nyu.edu/econ/cbp/

8) Dicennial Census and American Community Survey (2017-2021, last updated January 2023).Estimated percent of people with at least a Bachelor's degree, between 2018-2022. https://www.policymap.com/.https://data-census-gov.proxy.library.nyu.edu/cedsci/

9) Dicennial Census and American Community Survey (2017-2021, last updated January 2023).Estimated income of a household, between 2018-2022. https://www.policymap.com/.https://data-census-gov.proxy.library.nyu.edu/cedsci/

10) Dicennial Census and American Community Survey (2017-2021, last updated January 2023).Estimated median age of all people, between 2018-2022. https://www.policymap.com/.https://data-census-gov.proxy.library.nyu.edu/cedsci/

11) Horowitz, J. M., Igielnik, R., & Kochhar, R. (2020, January 9). Trends in Income and Wealth Inequality. Pew Research Center. https://www.pewresearch.org/social-trends/2020/01/09/trends-in-income-and-wealth-inequality/

12) Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., & Wernerfelt, N. (2020, August 1). Social capital I: measurement and associations with economic mobility. Nature. https://www.nature.com/articles/s41586-022-04996-4

13) Mincer, J. (1958). Investment in human capital and personal income distribution. Journal of Political Economy, 66(4), 281-302.

14) Miller, C. C., Katz, J., Paris, F., & Bhatia, A. (2022, August 1). Vast new study shows a key to reducing poverty: More friendships between rich and poor. The New York Times. www.nytimes.com/interactive/2022/08/01/upshot/rich-poor-friendships.html

15) Occhipinti, J.-A., Buchanan, J., Hynes, W., Eyre, H. A., Tran, K., Song, Y. J. C., Fels, A., Skinner, A., Rosenberg, S., Prodan, A., Ujdur, G., Natsky, A. N., Henderson, T., Hasudungan, R., & Hickie, I. B. (2023). Estimating the mental wealth of nations: valuing social production and investment. Nature Mental Health, 6(4), 247-253. https://www.nature.com/articles/s44220-023-00044-w

16) Kenton, W. (2022, November 27). What is social capital? Definition, types, and examples. Investopedia. Retrieved from https://www.investopedia.com/terms/s/socialcapital.asp

17) U.S. Department of Commerce. (2023, June 15). Geographic inequality on the rise in the US. U.S. Department of Commerce. https://www.commerce.gov/news/blog/2023/06/geographic-inequality-rise-us

18) Hanauer, N. (2019, July). Better schools won't fix America. The Atlantic. https://www.theatlantic.com/magazine/archive/2019/07/education-isnt-enough/590611/

19) Lewis, H. (2012, February 29). Museums can help address inequality. Museums Journal. https://www.museumsassociation.org/museums-journal/opinion/2012/02/01032012-museums-can-help-address-inequality/#

20) Winston, M. (2022). Economic inequality: The role of libraries in fostering upward mobility and social justice. Journal of Information Ethics, 31(2), 38-47. https://www.proquest.com/docview/2756262968/fulltext/32D2FDBEB0C64A7CPQ/1?sourcetype=Scholarly%20Journals

21) McNeill, H. (2020, July 31). The fitness divide. *GRC Global Group Insights*. https://insights.grcglobalgroup.com/the-fitness-divide/

22) USC Bedrosian Center. (2019, January 28). Snap peas and socioeconomic inequality: The role of farmers markets in addressing food insecurity. USC Bedrosian Center.

https://bedrosian.usc.edu/snap-peas-and-socioeconomic-inequality-the-role-of-farmers-markets-in-addressing-food-insecurity/

23) Florida, R. (2019, January 22). The great divide in how Americans commute to work. Bloomberg.https://www.bloomberg.com/news/articles/2019-01-22/how-americans-commute-to-work-in-maps

24) Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." Journal of Economic Perspectives, 27 (3): 79-102. https://www.aeaweb.org/articles?id=10.1257/jep.27.3.79

25) List, J. A., Momeni, F., & Zenou, Y. (2020, December). The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods. National Bureau of Economic Research (NBER). https://www.nber.org/papers/w28283

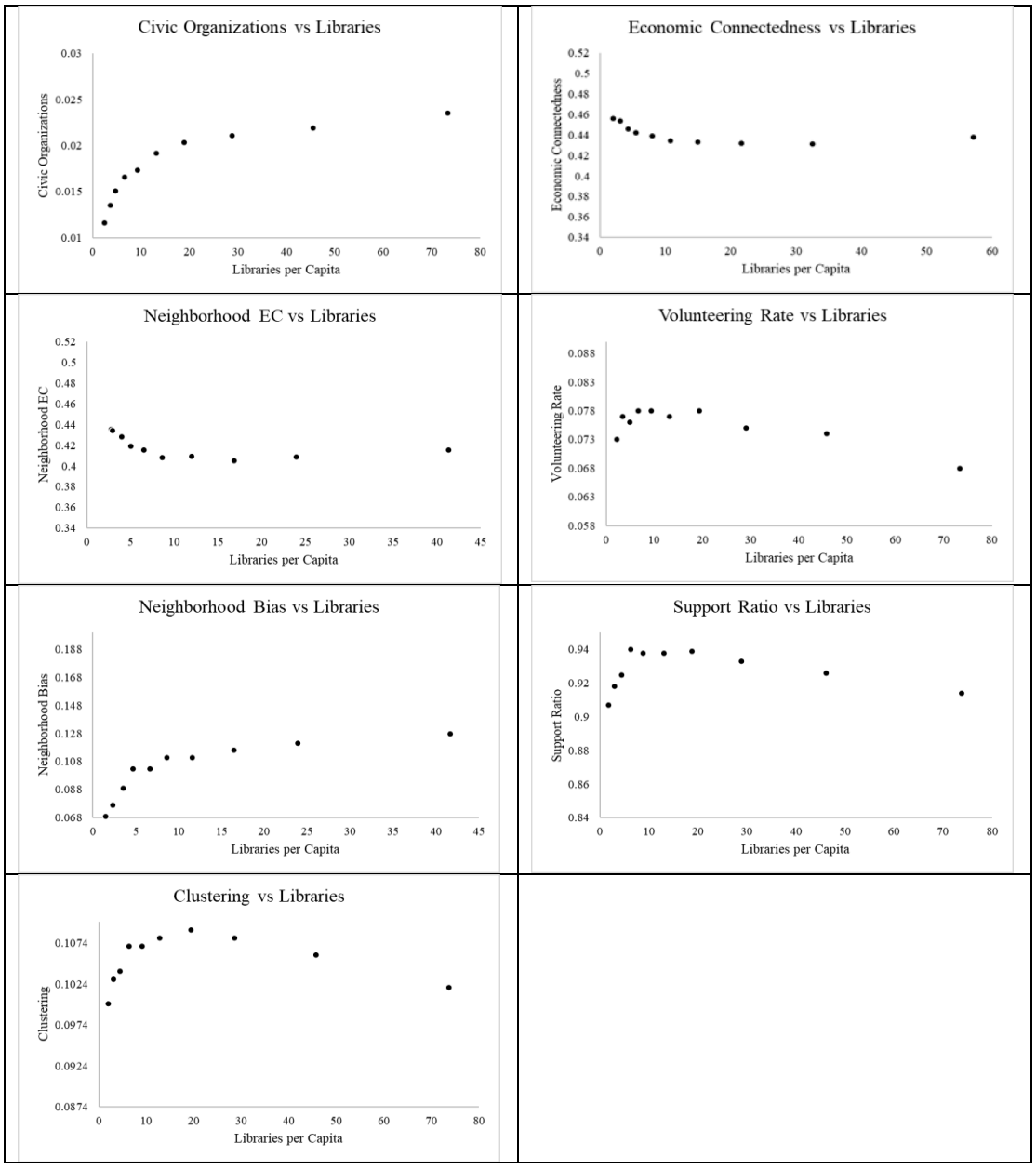
26) Kasakove, S., & Gebeloff, R. (2022, July 6). The shrinking of the middle-class neighborhood. The New York Times. https://www.nytimes.com/2022/07/06/us/economic-segregation-income.html

27) Opportunity Insights. (2022). About the Social Capital Atlas [PDF]. Retrieved from https://opportunityinsights.org/wp-content/uploads/2022/07/socialcapital_about.pdf

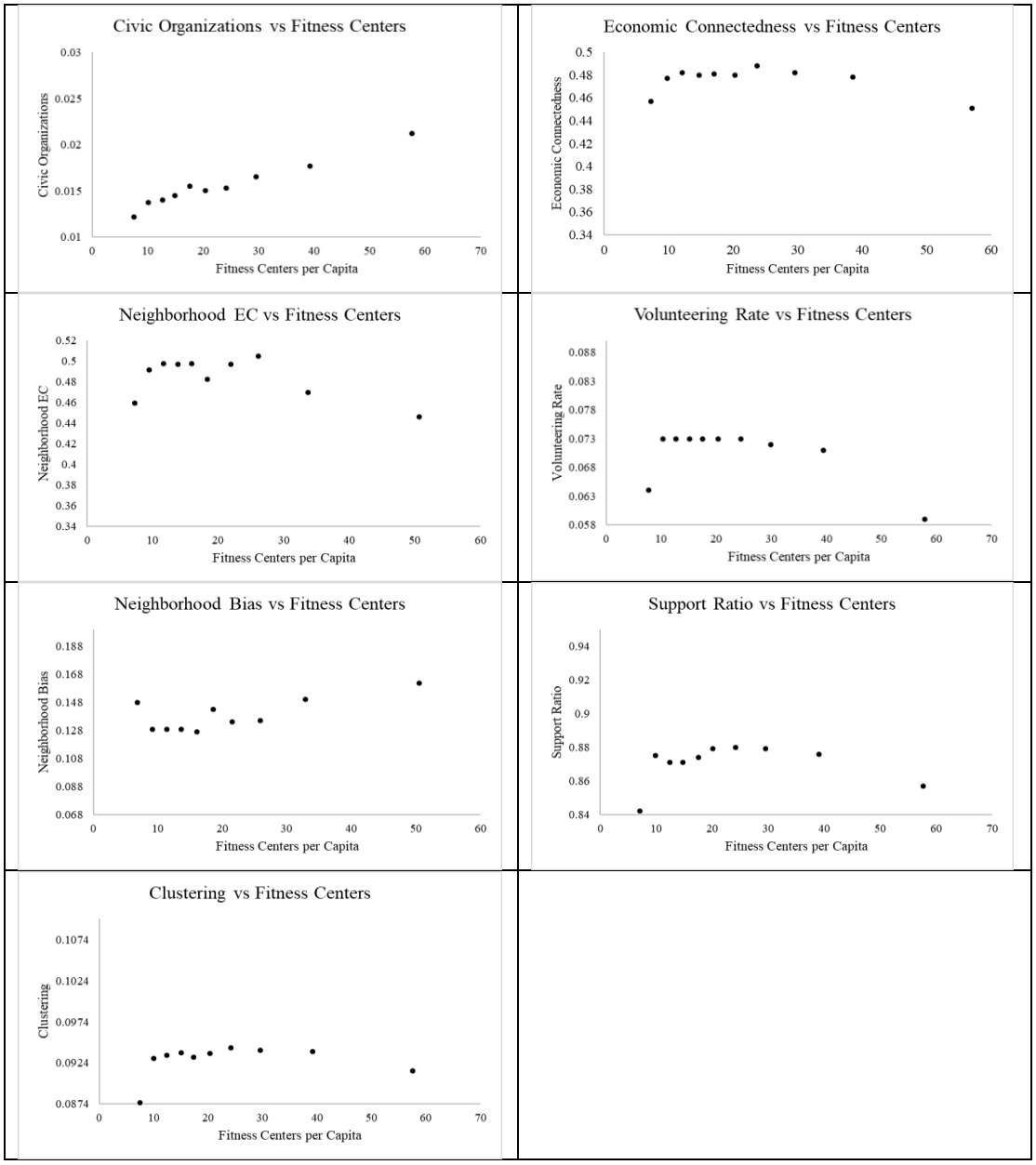
VI. Appendix

In this appendix, the binned scatterplots are arranged and displayed by the various explanatory variables (libraries per capita, farmers markers per capita, etc.) in order to determine the associations between the public infrastructure and the social capital measures.

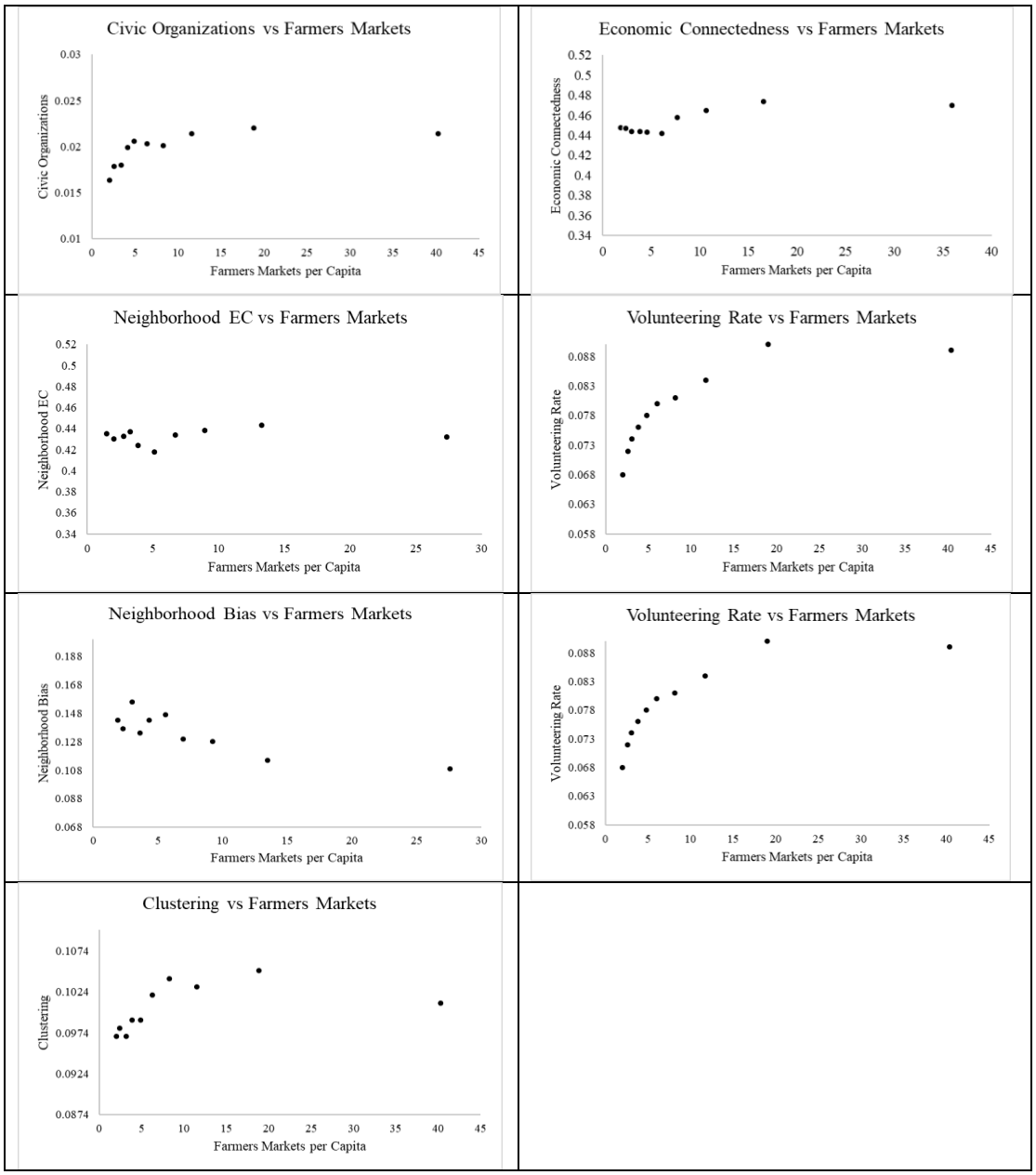
Appendix A1: The association between libraries per capita and the social capital measures



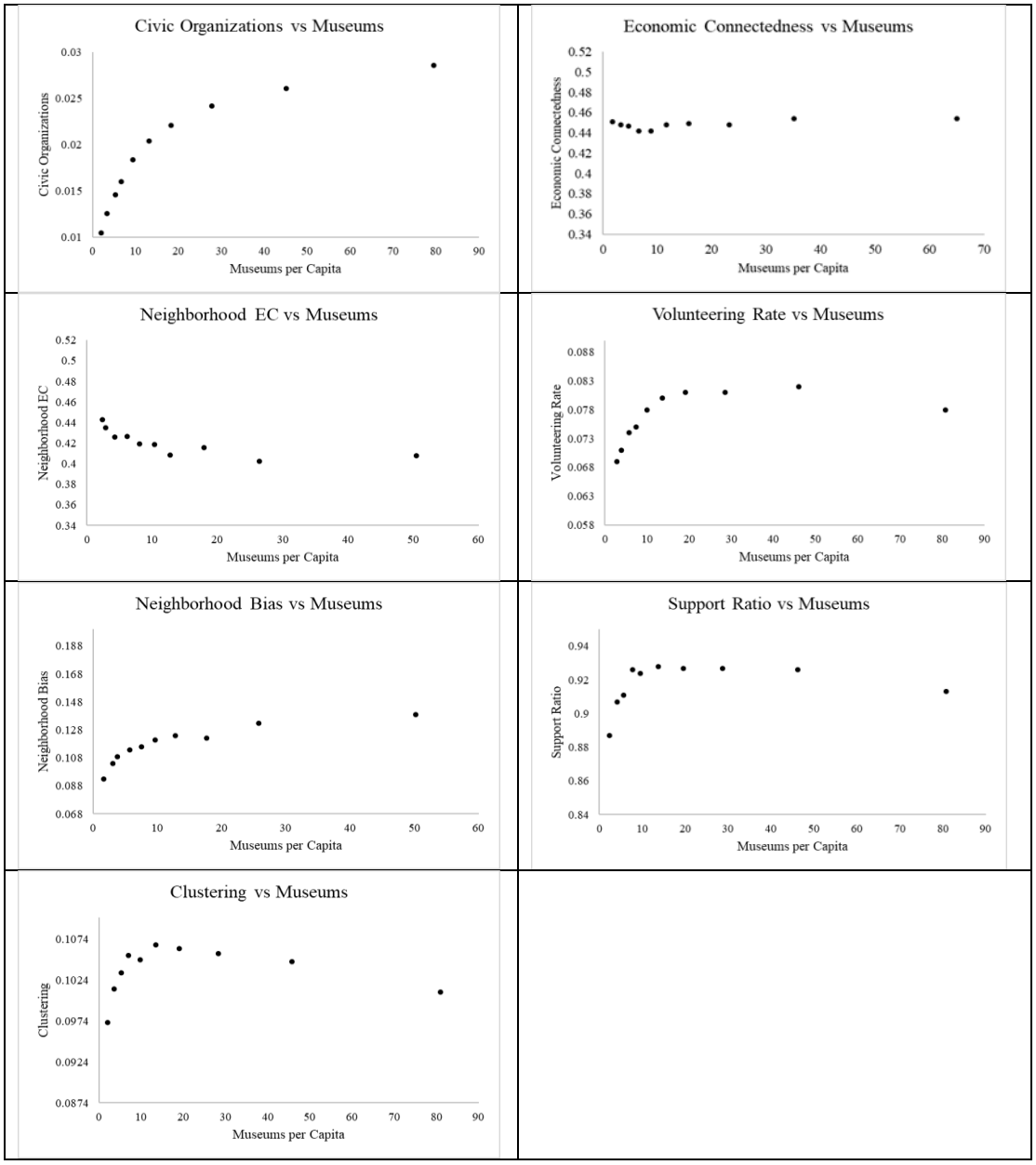
Appendix A2: The association between fitness centers per capita and the social capital measures



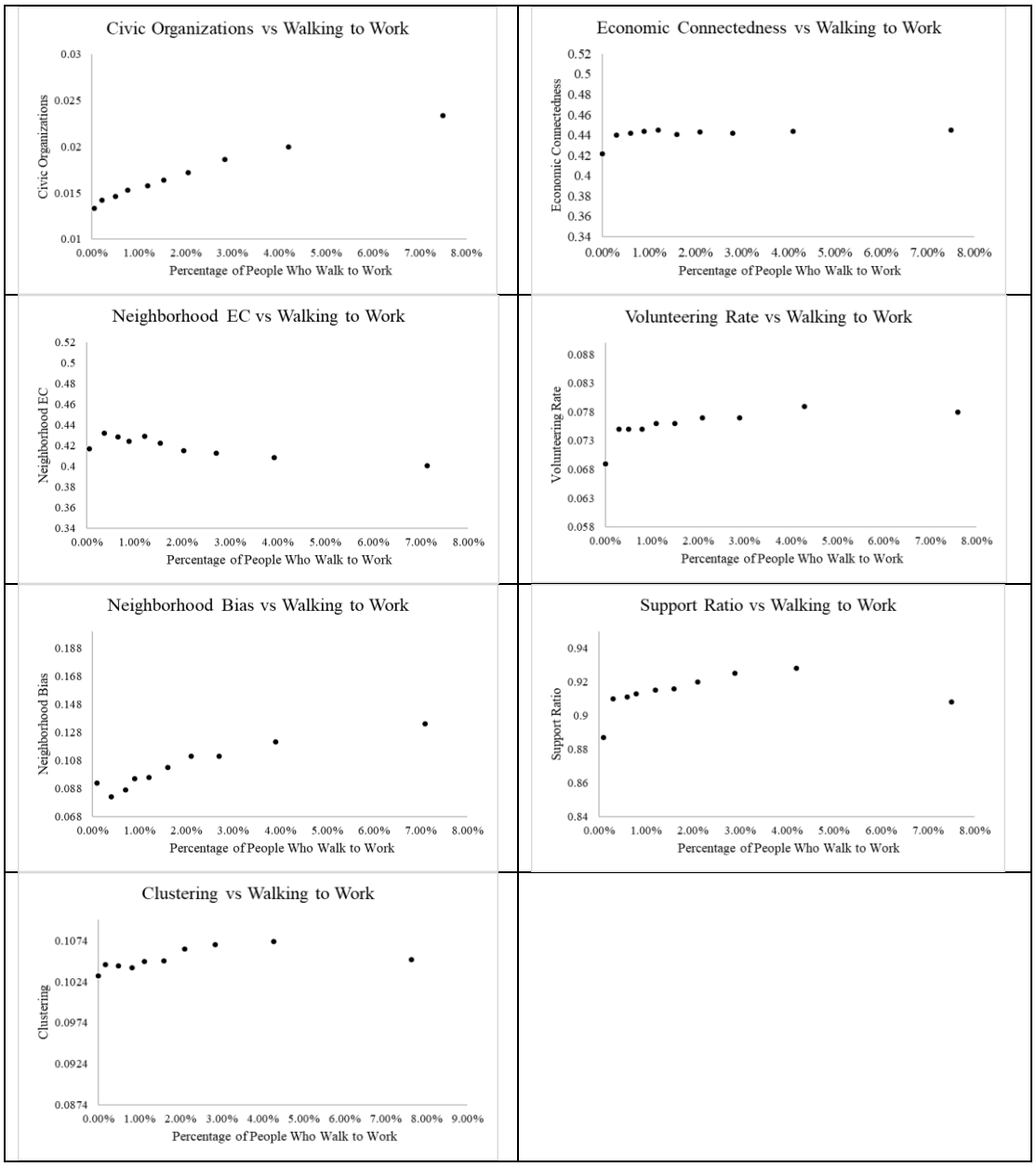
Appendix A3: The association between farmers markets per capita and the social capital measures



Appendix A4: The association between museums per capita and the social capital measures



Appendix A5: The association between the percentage of people who walk to work in a zip code and the social capital measures



Appendix A6: The association between the average number of vehicles per household and the social capital measures