The Effects of Transportation Access on Partisan Clustering:

Does transit access in American metropolitan regions retain partisan diversity?

Ryan Joseph Mauro

MT Supervisor: Enrique Hernández Pérez

Treball de Recerca

Master Thesis presented with a view to obtaining the degree of Master in Political Science,
Universitat Autónoma de Barcelona

Barcelona

July 2017

ABSTRACT

It is often argued whether the American electorate is in a period of polarization (Abramowitz 2010), partisan sorting (Fiorina & Abrams 2009), or geographic clustering; and what might be the causes of these observations. Using election results from the 2012 presidential election and transit ridership data at the county level this article argues for the structural causes of partisan spatial segregation and seeks to show how variation in transportation access can influence partisan clustering. This study does not find support for the ability of public transit access to decrease partisan clustering, but does find some support for the hypothesis that an increase in transportation access correlates to an increase in Democratic partisan clustering.

Keywords: Electoral Behavior, Geographic Sorting, Partisan Clustering, Polarization, Mass Transit, Urban Planning, Geospatial design impacts, 2012 US General Election

RESUM EXPLICATIU

Sovint es discuteix si l'electorat nord-americà es troba en un període de polarització (Abramowitz 2010), classificació partidista (Fiorina & Abrams 2009), o agrupació geogràfica; I quines poden ser les causes d'aquestes observacions. L'ús de resultats electorals a partir de les eleccions presidencials de 2012 i les dades del pilot de trànsit a nivell del comtat argumenta les causes estructurals de la segregació espacial partidista i pretén mostrar com la variació en l'accés al transport pot influir en el clúster partidista. Aquest estudi no troba suport per a la capacitat d'accés del trànsit públic per reduir el clúster partidista, però sí que dóna suport a la hipòtesi que un augment de l'accés al transport es correlaciona amb un augment del clúster Partit Demòcrata.

Paraules Clau: comportament electoral, ordenació geogràfica, agrupació partidista, polarització, trànsit massiu, planificació urbana, impactes del disseny geoespacial, elecció general dels EUA 2012

TABLE OF CONTENTS

Index of Figures, Tables & Equations:

I. Introduction

Many scholars have observed the clustering of Americans into homogenous partisan communities (Bishop 2009; Gimpel 2013; Lang & Pearson-Merkowitz 2015). This trend, similar to the polarization of the electorate, is seen as troublesome within democratic political discourse as legislative districts become homogenous clusters of partisans and ultimately unrepresentative. Of interest to scholars and policy makers might be how to address this phenomenon. Some scholars have found evidence to suggest that the phenomenon is caused in part by structural attributes such as infrastructure design (Nall 2015). By comparing the extent to which metropolitan regions throughout the US have or have not developed public transit systems to connect peripheral areas to urban centers we can observe to what extent access to public transportation has stymied the geographic sorting of the American electorate into partisan communities.

Transit systems cost a lot of money and thus require a significant amount of political will. States, cities, and counties must often cooperate with federal politicians and officials in order to raise the required revenue. Mass transit systems have been advocated for a variety of reasons: to combat the carbon emissions of personal vehicles, to decrease commute costs to and from job centers, to increase physical activity for health systems burdened by obesity, and to reclaim valuable public and private space that is often reserved for parking and personal vehicle use. These aspirations are mutually beneficial, but if it is found that public transportation access increases the vote share of one party over another then politicians will have incentive to oppose public transit expansion into their districts in order to keep their seats safe from competition. These incentives would be expected to contribute to a feedback loop of ever increasing partisan clustering in a system like that of the US where politicians choose their voters through redistricting and gerrymandering. This study does not find support for the ability of public transit use to decrease partisan clustering, but instead finds some support for the hypothesis that increases in transportation access correlate to an increase in Democratic partisan clustering. Additionally, this study finds support for the claim made by social interaction theory that increases in space for social interaction cause political participation to have a tolerance inducing effect as observed by reduced partisan spatial segregation.

Measuring partisan clustering is often measured by either comparing the vote differential or vote share of the Democratic and Republican parties at the lowest geographical level available, with greater differentials suggesting greater sorting. This study will measure partisan clustering by using the 2012 presidential election vote share in calculating the dissimilarity value of each county in relation to the state—a measure commonly used to measure segregation by comparing a smaller geographic unit's demographics to an aggregated unit's demographics, but which has also been used to measure partisan clustering (Walker 2013; Wong 2003). This dissimilarity value will show how clustered a county is in comparison to the state in which it resides. A check on the robustness of this variable will be done in the appendices of this paper to compare the results to national vote share dissimilarity and vote differentials. By applying intergroup contact theory, urban theory, social capital theory, and class voting theory this research will seek to explain how transportation access might contribute to partisan clustering.

This article will begin by laying out the theoretical framework for the analysis of partisan clustering in the USA, and then hypothesize how it relates to transportation access. The following section will then summarize the methods and data that will be used, concluding with an analysis of the data, and a discussion of the primary findings.

II. Background

2.1. Is there Partisan Clustering in the US?

Partisan clustering, known in other terms as *geographic sorting* or *partisan spatial segregation*, is the phenomenon where people live in homogenous communities that reflect their beliefs and political ideology. This occurrence has great importance in the American electoral system, as it would in other electoral systems of majoritarian single member districts, because of the ability to draw district lines so that a community can lie within either a competitive or non-competitive district. This phenomenon is also not restricted to the US for countries all around the world—and in particular Britain and its former colonies—demonstrate that left wing partisans are much more concentrated in small urbanized spaces then are right wing partisans (Rodden 2010). In the US as a cause of the redistricting process nearly 87% of US House seats in the 2018 elections are already predicted to be safe from competition (The Cook Report 2017). Researchers disagree about whether these communities are sought out by individuals seeking competitive or safe partisan districts—and thus created by the intention of individual actors to self-segregate—, if they are caused by structural factors that favor certain groups of people living in certain areas, or if they are created by politicians themselves in the redistricting process.

The topic of partisan clustering in the US received a boost of attention after the publication of Bishop's *The Big Sort* (2009). In it the author argues that the intentional migration of individuals into communities with similar political beliefs explains the reduction seen in competitive districts in the past few decades. Since its publication Bishop's work has spawned many empirical studies primarily on the hypothesis that partisan migration to likeminded communities is the cause of increased partisan clustering. This is an intuitive argument based on the assumption that voters supporting the minority party in a district will seek out representation in districts where their party has a chance of winning and representing them, however this assumes that most of the migration contributing to the phenomenon is from voters with high political knowledge who place high importance on politics. Some have found empirical evidence in support of this partisan migration hypothesis (Motyl, et al 2012; Gimpel, et al 2013; Sussel 2013; Lang & Pearson-Merkowitz 2015), while others have found evidence against it suggesting instead that relocation creates partisan clusters because those who identify with one party place higher importance on certain structural attributes of a place than do those who identify with another party (Mummolo & Nall 2017). Nonetheless, the observations of the American electorate as sorted geographically appear to be unchallenged by both supporters and opponents of the partisan migration hypothesis and thus still open to causal explanation. This research will focus on structural explanations for the observance of partisan clustering.

In the study of partisan clustering the assumption might come across that communities that vote in unison are somehow undesirable. Ideas and ideologies are not required equal representation in democratic republics but rather it is each voter and their ideas which are. Political diversity then in itself is not necessarily a goal nor a benefit, but an expected outcome of some theories of voting when community demographics are diverse. If class and social cleavages are not reflected in the vote share of a diverse community then a more likely explanation might be due to the party system in question not representing traditional cleavages. In the case of the US this is likely part of the explanation because of the presence of only two major parties both of which could be considered catch-all parties, and no traditional class based parties exist. Nonetheless social cleavages are still used with accuracy to explain the vote, with middle and high income, Christian, rural and suburban white voters consistently being shown to identify with the Republican party, while educated, low income, urban, and minority voters shown to identify with the Democratic party (Manza & Brooks 1999). With this understanding of class voting and social cleavage voting theories we can see that there are cleavages that each party represents and would expect to see their support reflected in the vote proportionate to the presence of these cleavages. If class voting theory is applicable in this context then we would expect to observe a correlation between the presence of services used by one class more than another—public transit, and affordable housing, for instance—and support for their "class party".

2.2. Theoretical Framework

A mechanism to explain the political divisions observed between communities with access to public transportation and those without is that public transit increases the Social Interaction Potential (SIP) of a community. SIP is a measure of a community based on the presence of spatial design features that foster interaction with others (Farber & Li 2013). In this metric communities with low population density, few pedestrian walkways and heavy reliance on single occupancy vehicle transport would have a low SIP score; while communities with high population density, generous pedestrian walkways and high public transit use will have higher SIP scores. Given that public transportation is an indicator for the metric its existence in a community would likely increase a SIP score even in a community with low density like America's exurbs (Farber & Li 2013; Currie & Stanley 2003). Little research using SIP scores has been done but measuring its correlation to that of the geographic distribution of partisan sorting can be an interesting test for its application as a mechanism.

Public transportation access as a conduit of social interaction then acts as a bridge for social capital to cross community lines and strengthen ties and contribute to a beneficial mutual understanding of the other, whereby other forms of social capital derived from insular communities might increase social exclusion of others (Currie & Stanley 2003). Communities without public transportation access would then be expected to have low turnout rates where there is a presence of socially excluded outsiders—such as non-partisans—who are not in a position to migrate to locations more similar to them. This would then be compounded by the nature of the first past the post single district electoral system and gerrymandering for US house seats, as well as the construct of non-competitive winner take all states for the presidential election.

The effect that social interaction could have on partisanship is still contested. Social capital theories might argue that the absence of social interaction is alienating, disempowering, and feeds into divergence of political opinions through self-selected publics (Putnam 2001; Stolle & Rochen 1998). This would suggest that with less public transportation—and thereby less social interaction—communities would either be more polarized or homogenous through self-selection. Other theories on how social interaction can affect partisanship have found context dependent results showing that views are more likely to be moderated when confronted by dissonant opinions more so when those views are in a minority position to a majority dissonant position (Huckfeldt 1987). Other theories, such as Contact Theory and Urban Theory, still point to and even advocate for face-to-face intergroup interaction as a way to create tolerant communities (Wessel 2009; Fainstein 2005). In areas with high political, economic, and ethnic diversity public transit's ability to increase social interaction by forcing commuters and travelers to confront inhabitants outside of their isolated and segregated communities rather than driving through them should then effect transit users' political views of non-partisans. However increases of low-quality interactions with others outside one's community could also have the effect of decreasing social capital and thus one might expect to see even greater geographic clustering (Putnam 2007).

Allport's "intergroup contact theory" specifies that in order for contact to have a beneficial or tolerance inducing effect contacts need to have "equal status", "common goals", and "support from authorities or custom" (Pettigrew 1998, p.67). Recent research suggests that support from an authority figure for the mediating behavior is even more important in today's current context due to increased polarization and partisan sorting making norms formation a highly partisan issue (Dyck & Pearson-Merkowitz 2014). Assuming that both parties in the US have equal ability to attract tolerant independent voters with vocal mediating figureheads then we would expect to find of these results that transportation access correlates with competitive races with a low vote differential. Thus, deriving from the theories related to social capital and intergroup contact the first hypothesis of this research will be:

> **H1**: Communities with greater access to public transportation will have lower partisan clustering, having a lower vote differential.

Due to social capital theory's findings on how interaction might increase political participation, there would appear to be an interactive effect between transit access and political participation. Higher levels of transit would increase social capital which would decrease partisan clustering, in relation to effective social interaction boosting political participation. An interactive model would then be necessary to test this hypothesis.

> **H2.1 :** The marginal effect of *public transportation access* on *partisan clustering* will be positive when values of *political participation* are at their lowest levels, and negative when values of *political participation* are at their highest levels.

This interactive hypothesis comes from the assumptions that with higher transit use the social interaction experienced on transit will increase political participation through social capital building having an overall marginal effect to decrease partisan clustering. The other part of this hypothesis assumes that higher levels of political participation in a community will interact with increased transit use to have a marginal effect in the direction of reducing partisan clustering. This naturally then produces a second prediction on how these variables might interact with transportation access acting as the intervening variable and political participation as the main independent variable:[1]

> **H2.2**: The marginal effect of *political participation* on *partisan clustering* will be positive when *transportation access* is at its lowest values, and negative when *transportation access* is at its highest values.

If instead one party more than another represents tolerance and diversity then that party would be expected to benefit from a large vote differential where transit access is highest. Assuming, based on sorting or elite polarization theories, that the GOP has become the party of conservative ideology—which is one "associated with stereotyping prejudice, intolerance, and hostility toward a wide variety of outgroups"—then we would instead expect to find evidence in support of a hypothesis that the differential would increase to the benefit of the Democratic party where public transit is accessible (Jost et al. 2009, p.325). A theoretical framework to support a hypothesis where transportation access increases Democratic party clustering would then need to be developed.

Before developing the theoretical framework behind how transit access might affect Democratic party clustering it is important to state that county and state governments have an interest in maintaining unity across rural-urban divides and they do this by connecting

[1] Testing an interactive hypothesis from both sides of the interaction allows for a clearer understanding of how an interaction behaves. Refer to: (Berry, Golder & Milton 2012)

residents of all areas and income to employment and opportunity through infrastructure projects so as not to encourage breakaway states. Car ownership and highway projects are one way used to provide access to employment opportunities, but the costs make it out of reach for many. Nall (2015) found that the funding of the Interstate Highway program created suburban communities that favored wealthier residents with the ability to afford automobile ownership expenses, leading to Republican party clustering in communities with freeway access. On the other hand Glaeser, et al (2003) found that access to public transportation made living in central urban areas affordable for the poor, even when housing was not as affordable. In terms of causal mechanisms Nall's (2015) and Glaeser et al's (2003) mechanism can be summed up as greater income diversity allows for greater political diversity and by extension more competition in suburban communities. Others have also noted the potential of public transit to increase the wealth of those communities with access to it, which could contradict or complicate the hypotheses presented here (Banerjee et al 2012).

In 2017 Mummolo and Nall released a study which found support for the structural explanation for partisan clustering. In this study the authors surveyed self-identified Democrats and Republicans and asked them to rate how important features of a given community would be in deciding whether or not to move there. There were some features which obviously Democrats and Republicans placed equal importance on—such as a community's safety, access to employment, climate, proximity to family, and school quality—, but there were other community features that one party seemed to cherish much more than the other when choosing a place in which to relocate. Democratic survey respondents were found to place greater importance on a community's access to low-income services, neighborhood walkability, proximity to metropolitan areas, and public transit quality. Republican respondents on the other hand ranked community wealth, taxes, and big houses much higher in importance than Democratic respondents. The explicit partisan migration question as to whether it mattered that their neighbors share their political views was also asked in the survey and was ranked equally low by both Republican and Democratic respondents (Mummolo & Nall 2017). The purpose of testing for the geographic clustering of partisans in this way is that it tries to take into account that clustering might not be intentional but accidental due to partisans selecting communities to live in that non-partisans wouldn't care for. This method however has lower internal validity because it does not explicitly stick to a minimalist understanding of agent driven partisan clustering, but can be useful in exploring the factors that incidentally cause and replicate the phenomenon. If their findings are accurate then we should expect to find more Democratic party voters where public transit accessibility and ridership is high. This insight then leads into the third hypothesis of this study:

> **H3**: Communities with greater access to public transportation will be observed to have greater Democratic party clustering

However it must also be noted that although there is a noticeable difference in how important this issue is between partisans of each respective party, a lower proportion of the sample population finds these issues important for community selection overall when compared to other community characteristics like location, orderliness neighborhood income, and government (Mummolo & Nall 2017). This should cause for a reduced effect overall by transit access on partisan clustering. In order to accurately measure this effect other variables that also diverge in importance for community selection along partisan lines should also be taken into account and controlled for.

There is also the possibility that public transit access increases housing costs and real estate values (Banerjee et al 2012) which would then be expected to decrease the effect of transportation access on class voting and class clustering because Republicans place more importance to real estate prices and neighborhood wealth than do Democrats (Mummolo & Nall 2017). This effect would then again be mediated according to the extent to which having higher housing costs is offset by transportation costs and quality (Glaeser et al 2008). This theory combined with class theory and social cleavage voting theories suggests that there might be an interactive relationship between housing and transit costs and transit access. An interactive hypothesis would then need to be generated.

> **H4.1**: The marginal effect of *transportation access* on *Democratic partisan clustering* will be positive when *housing and transit costs* are at their lowest values, and negative when *housing and transit costs* are at their highest values.

The interactive hypothesis above is based on the assumption that the affordability of housing and transit options will mediate the effect transportation access has on Democratic party segregation. Low affordability would then likely create a negative marginal effect from transit access increases on Democratic partisan clustering. This hypothesis then produces a second prediction on how these variables might interact with the variable's roles in bringing about an effect switched:[2]

> **H4.2**: The marginal effect of *housing and transit costs* on *Democratic partisan clustering* will be negative across all values of *transportation access.* When *transportation access* is at its lowest values the marginal effect will be strongest and most negative, and when *transportation access* is at its highest values the marginal effect will be at its weakest.

III. Research Design

3.1. Unit Of Analysis

The Unit of Analysis will be all of the 1,800 US counties that are located in the Census Bureau's Core Based Statistical Areas which combines together counties that have strong economic and cultural links to a core metropolitan area. Using counties as the unit of observation was largely influenced by the precedent set in Nall's 2015 study on highway access and partisan clustering. The counties selected show great variation in the presence of the independent variable and the dependent variable so could be said to be using the method of difference in case selection, however the selection was primarily driven by data availability for the independent variables. The Housing and Transportation Affordability Index (H+T Index)—the source of data for the independent variables in this study—provides information for these counties (The Center for Neighborhood Technology 2016). By selecting counties in CBSAs there should be an expected amount of commuting from central and outlying counties to metropolitan centers which should be reflected in the public transportation access data. As the county level is smaller than both the state and CBSA levels this should help to show any spatial differences that might vary with partisan clustering.

The county level as a unit of analysis is not the most ideal due to the wide variation between counties on the basis of size and population. For example, the least populous is McPherson County, Nebraska with just 382 estimated inhabitants in 2013, while the most populous being Los Angeles County, California with over 9 million inhabitants. By land area there is also significant variation: from Falls Church County, Virginia with just 5.14 square

---

[2]For more on the logic behind making two prediction for each interaction refer to: (Berry, Golder & Milton 2012)

kilometers of land, to San Bernardino, California with nearly 52,000 square kilometers of land. These large variations in the characteristics of the unit of analysis can be cause for some major problems. Much variation that could be observed within these large counties will go unnoticed and lost into the aggregate means and medians of the county as a whole, being blotted out by the more densely populated centers of the county. At this point there is also a regional difference in county land size, with counties in the eastern half of the US having much smaller land area than those in the western half of the US. This difference is likely a cause of the West's late development of a county system, lower population densities, and topographic differences caused by the presence of mountains and deserts. The result of this regional difference means that observations from eastern counties will be more reflective of the population due to less variability possible with a smaller population and less land area to provide transit services, while observations from the western states will be less reflective due to increased variability within the county.

Another issue with using counties as the unit of analysis is the crisscross of legislative districts between and within counties. These governing bodies have significant control over spending decisions that might impact the presence of accessible transit in their counties, as well as which voters they want in their districts through redistricting and gerrymandering. The variation between districts will go largely unobserved in using county level data. However, a benefit remains by using county data over districts in that legislative districts are constantly changing shape to the extent that a cause and effect of transit access on partisan clustering could not be attributed to district legislative representation because public transit deals could have been made months or years before when that county was part of an entirely different district with different legislative representation. So, by sticking to counties for observation some historical linearity can at least be inferred through the data even when the analysis is of just one year because of the consistent boundaries of the county.

Ideally this research would be done using an even smaller unit of analysis, which does happen to be available for the independent variable. However, data at lower units are not freely available for all cases under consideration. Alternatively, smaller units could be observed if one restricts their scope to a handful of case studies, whereby there is no doubt excellent data is being published at the local level for local political office, but then one encounters the particular issues of observing state and local elections where voter turnout is greatly reduced to the point that it might skew interpretation of partisan clustering (Hajnal &

Lewis 2003; Marschall et al 2011). For these reasons, this study will use the county level for analysis.

3.2. The Dependent Variables: Partisan Clustering

The operationalization of partisan clustering is more complicated than that of the independent variables. Measuring the geographic clustering of partisans would ideally be done by attaining the political attitudes at a level as small as neighborhoods. The level of political homogeneity of communities would then ideally be measured by surveys of political ideology along a left-right scale, or level of partisan identification. This would best take into account and hold for the regional differences observed in the US, where conservatives voting for a Republican candidate in one part of the state might be liberals who vote for a Democratic candidate in another (Feinberg et al 2017). In this case the American National Elections survey would be ideal by selecting responses only from those zip codes of interest. However, the ANES survey along with the American Community Survey and others like it cannot be used either because there are too few responses per unit of analysis, or because of data restrictions on geographic indicators of respondents. Voter registration data would be another possibility to measure partisan clustering, but many states do not require voters to register with a party so would not be comparable across all units. Without sufficient resources to dispatch a survey nor current data availability at such a level of analysis election returns become the next best option for operationalizing the dependent variable.

This then leads to the next issue as to whether partisan clustering would be measured best with presidential election returns, or congressional election returns. Using the presidential election results at the county level (McGovern & Larson 2017) avoids the problem of low voter turnout as caused by gerrymandered districts and second-order elections, which should then give a less biased picture of the partisan makeup of a county. The two common methods of measuring partisan clustering are either through using the vote share of a single party in a presidential election, or the vote differential between the two parties. The vote differential has the benefit of describing the direction of partisan clustering through negative and positive values. This measure would be calculated by subtracting the Republican party's percentage of the vote share from the Democratic party's percentage of the vote share. For example, a 50-50 split in the 2012 vote for president in a county would mean a vote differential of 0, suggesting that there is no partisan sorting in that county. On the other hand, a county with 80 percent of the electorate voting for the Republicans and 25 percent voting for the Democratic candidate would then have a vote differential of -55, reflecting a potentially Republican party sorted county. This scale is one with 201 possible values, including zero, ranging from -100 to +100. For the first hypothesis this directional component of the differential is not desirable, and so the absolute value of the differential would need to be used instead, which would be calculated by taking the square root of the squared vote differential. In the previous example the county that had a -55 vote differential in the Republican direction would simply be considered as sorted to the value of 55 out of 100 in the test for Hypothesis 1.

Using the vote differential is useful but doesn't operationalize the concept of partisan clustering as well as we would like. In particular, the vote differential does not answer the question of whether a county is clustered in relation to the region, the state or the nation. This is an important distinction to make because a county that appears clustered by an absolute vote differential value of 40, for example, might be located in a state where the vote differential is also 40 meaning that the county is representative of the state and not clustered at all. By calculating the Dissimilarity value of the vote share, a measure often used to identify levels of segregation in a geographic area, some scholars have addressed this problem (Walker 2013). In this research the dissimilarity value will be calculated by taking the vote share of the Democratic party candidate in proportion to the vote share of that candidate at the state level and then subtracting that value by the vote share of the Republican candidate at the county level in proportion to the Republican vote share at the state level. Equations 1 and 2 below show the calculations for the dissimilarity directional values (1) and absolute values (2):

$$D = \frac{1}{2} \sum \left(\frac{a}{A} - \frac{b}{B}\right) \tag{1}$$

$$D = \frac{1}{2} \sum_i \left| \frac{a_i}{A} - \frac{b_i}{B} \right| \tag{2}$$

Where a_i and b_i are the partisan populations of the unit under analysis—measured by the county level vote share—and A and B are the partisan populations in the population as a whole—measured through the vote share at the state level (Walker 2013; Wong 2003). In order to check for the robustness of this indicator the dissimilarity value in relation to the national vote share can also be calculated for robustness checks at the end of the paper. This then provides us with an indicator that can be measured with 201 possible values from -100 to +100, much like the vote differential. This indicator (equation 1) can also be converted into an absolute value (the scalar value, equation 2) through the same calculation as used for the vote differential to test Hypothesis 1. The histograms in figure 3.2.1 show the distributions of the six dependent variables to be used in this research.

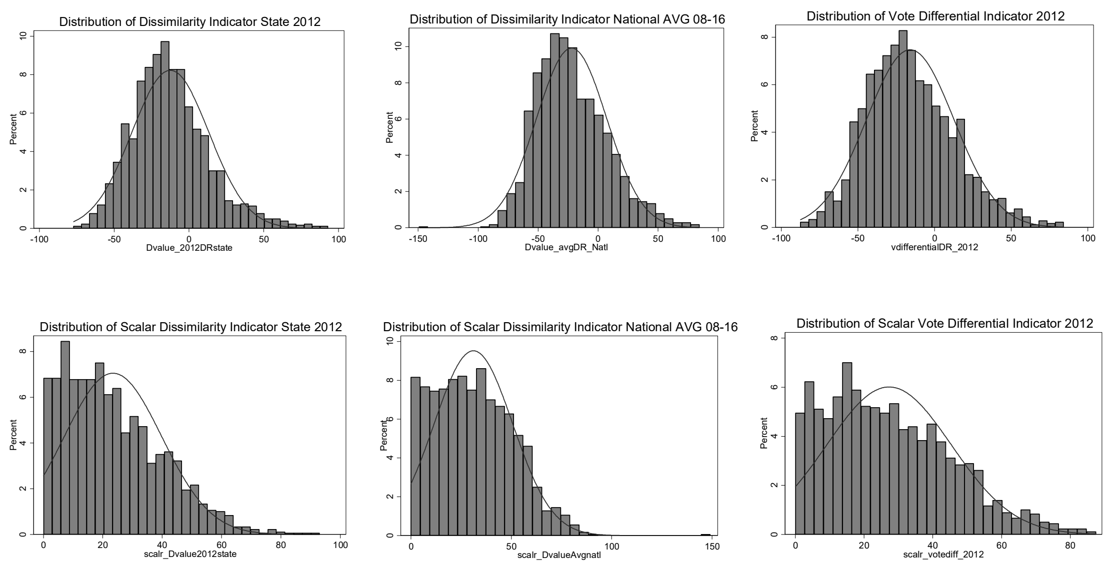

Figure 3.2.1 *histograms of the dependent variable indicators with the ideal normal distribution lines superimposed. The top 3 distributions are the vector indicators, and the bottom three are the scalar indicators. The left-most graphs are distributions from the Dissimilarity value calculated with state 2012 vote totals, the center two graphs are Dissimilarity values calculated from the national averaged vote totals for the 2008, 2012, and 2016 elections, while the two right-most graphs are distributions of the vote differential for 2012.*

The interpretation of the Dissimilarity score changes between the normal directional and absolute value versions. In the absolute value *D* represents the percent of a community's population that would need to relocate in order to proportionately reflect the partisan makeup of the larger community—in the case of the left-most indicators in Figure 3.2.1 the larger community is the state, while in the case of the center variable this is the country as a whole—but does not suggest which partisan population would need to relocate. In the vector version of the Dissimilarity score, *D* represents twice the percent of the population that would need to relocate in order to reflect the state or national partisan makeup, and then shows who would need to do the relocating: if *D* is negative than there are more Republican voters represented in the county than in the state or national electorate, while if *D* is positive than there are too many Democratic voters represented in the county than in the state or national electorate. These distributions of the dependent variable indicators also show a skew of counties towards Republican clustering.

Using the presidential election at the county level also runs into problems with the possibility of leadership effects biasing the observation of partisanship at the community level. One way to address this could be to average out the three most recent elections' vote differentials and vote shares at the county levels in order to get a more accurate representation of partisanship beyond leadership effects. Because the H+T Index data to be used for measuring the independent variable represents the single year of 2013 the presidential election results of 2008, 2012, and 2016 could then be used to ascertain the partisan makeup of a community. This average vote differential and dissimilarity value will then be used to check for the robustness of the variables at the end of the article.[3]

3.3.The Independent Variable: Public Transportation Access

Public transportation access can be described as: the affordability of transit, a community's proximity to transit stops and destinations, the time to get between city center and exurban periphery, and the frequency of transit. Finding a single variable that encompasses all these aspects can be difficult. Transit ridership will be used as the primary indicator for measuring transit accessibility, for it can be assumed that in order for there to be high ridership transit would have to be reachable, affordable, and have competitive trip times to other forms of transportation. Transit ridership does not offer as accurate a picture of transit accessibility as might multiple indicators, but it does have the benefit of being a succinct one in all indicator. Where ridership alone cannot explain variation other control variables may be introduced.

The data for transit ridership was compiled by the Center for Neighborhood Technology's H+T Index by aggregating data from the National Transit Database and the American Community Survey from their geographic coordinates (Center for Neighborhood Technology 2017). The group behind the dataset, the Center for Neighborhood Technology, is an nongovernmental organization that advocates for public transit and sustainable urban design. This dataset provides ridership information from each county's transportation departments into a measure of annual transit trips per year for the regional typical household. Another measure, which is arguably more relevant to this research, is the percent of all commute trips by public transit in a county. These two indicators, however, have significant skewness issues (see Figure 3.3.1):

[3] Refer to Appendix B for further explanations of the data used for the dependent variable and their means, standard deviations, and min/max values

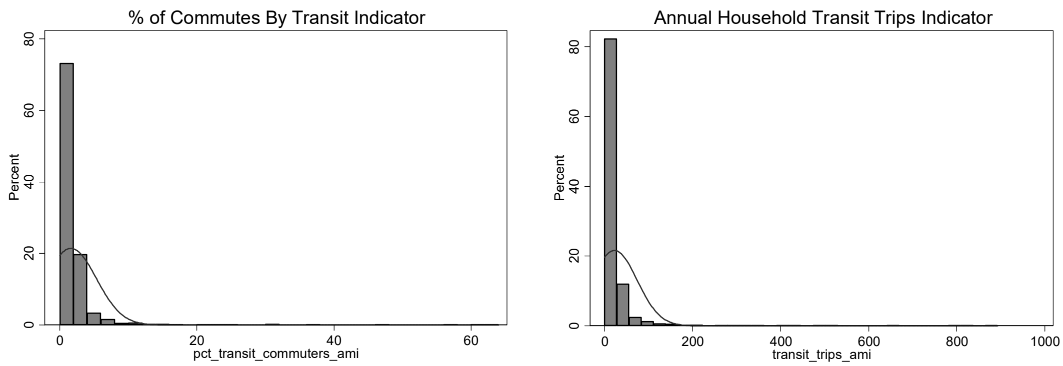

Figure 3.3.1 *The original data distribution from the Housing and Transportation affordability index shows a significant rightward skew; the ideal normal distribution line has been superimposed.*

To address this the independent variables have been logged transformed. This successfully created a log-normal distribution of the independent variables, however it should be cautioned that interpretation of this data has now been changed from a linear model to a linear-log model (Benoit 2011).

$$Y_i = \alpha + \beta_i \log X_i + \varepsilon_i$$

$$\log X + 1 = \log X + \log e = \log(eX) \qquad (3)$$

This means that in the following regressions that the coefficient for the independent variable, transportation access, will be expected to have a subsequent effect on the dependent variable—partisan clustering—when transportation access increases by *e* or by *172%*, instead of the normal regression interpretation of a 1 unit change (Benoit 2011).

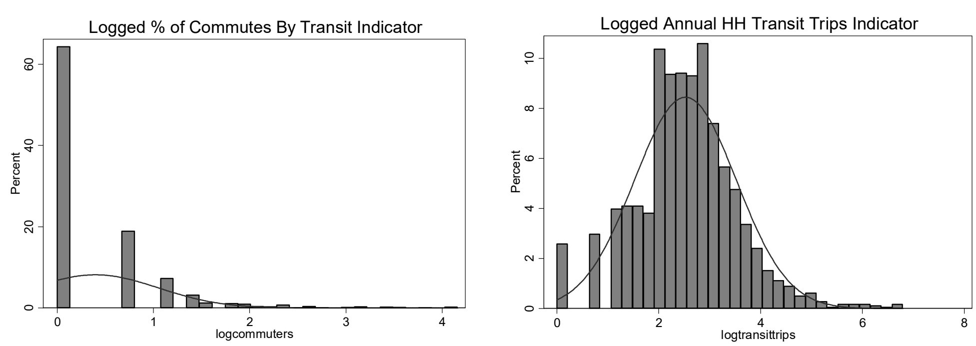

Figure 3.3.2 *Log transformation of the independent variable indicators with ideal normal distribution line superimposed (many observations were dropped due to zero values)*

This transformation was effective and worked nicely with following regressions, however that was only because it dropped a quarter of the data because its value was zero and the log of zero is incomputable. By using a log transformation and adding 1 to each value in the dataset for the independent variables these zero values can be retained and reflected in the dataset (se Figure 3.3.3):

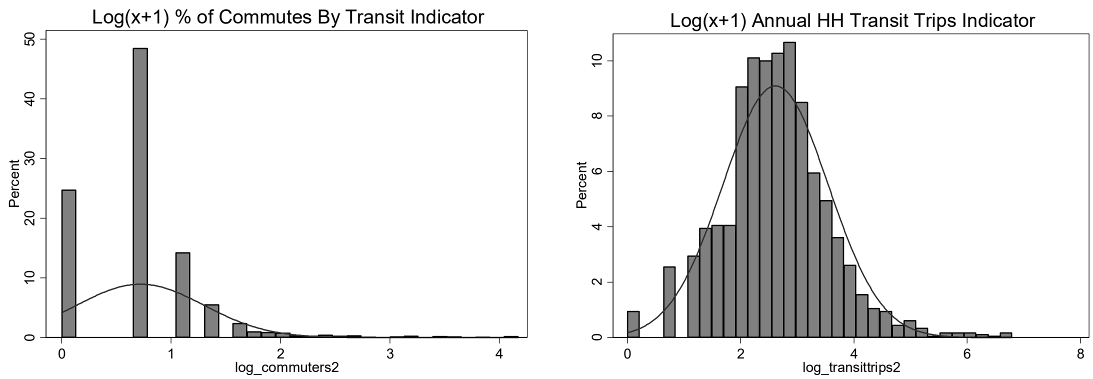

Figure 3.3.3 *log(x+1) transformations of the independent variable indicators with ideal normal distribution line superimposed.*

The log(x+1) variables will be used for the subsequent regression analyses in this article. With these variables in the model the interpretation of the coefficients changes somewhat from the prior linear-log model:

$$Y_i = \alpha + \beta \log(X_i+1) + \varepsilon_i$$

$$\log(X+1) + 1 = \log(X+1) + \log e = \log(e(X+1)) \quad (4)$$

This linear-log model will be interpreted as: the coefficient, β, represents a change in Y when $X+1$ is multiplied by e, or 2.72. In order to measure how much of a change in Y a one unit change in X makes the β coefficient would then need to be divided by 2.72 and then have 1 subtracted from it. The final transformations of both independent variables have a much more normal distribution than both the original and log transformed data sets, and so are apparently best suited for regression analysis.

In choosing the independent variable for measuring transportation accessibility there is a tradeoff. On one hand the best variable for understanding the effects on the dependent variable is the one that measures the percent of commutes that use public transit because of the simplicity of interpreting a one percent increase in transit use and a subsequent effect on partisan clustering. However, this variable has a much less normal distribution due in part to the original dataset rounding all the decimal percentages to the nearest whole number. The second independent variable that can be used does not suffer from this drawback and has a much more normal distribution, but is much more difficult to interpret because it is not a percentage but rather the county median income household's average number of public transit trips in a year. This means that in interpreting the coefficients in subsequent regression tables the increase in *x* that causes *y* is a meager 1 trip increase in a year. This has much less meaning, and so the percentage of commute trips taken by transit will be the primary independent variable. When necessary the annual transit trips variable will then be used as a robustness check on regressions made with the percentage of mass transit commute trips indicator.

The other independent variables that will be used are those which will test the interactive hypotheses of Hypotheses 2 and 4. In Hypothesis 2 the predicted interaction will be between transportation access and political participation. The way that political participation will be measured in this article is through the voter turnout for the county in the 2012 election. There are other ways that are perhaps more significant in representing levels of political participation for social capital theorists, but the easiest way for the unit of analysis chosen is through the voter turnout. The next choice is whether to measure the voter turnout of registered voters or eligible voters. Using registered voters as the total population of voters might unfortunately skew the data to make it seem that there is higher turnout than there actually is. For example, a registered voter population could be half that of the voting age population and all turnout to vote, which would make that county's turnout rate 100 percent when in actuality 50 percent of the electorate abstained. By measuring the voting age population turnout the theorized covariation between turnout and partisan clustering caused by gerrymandering, partisan and swing states can be observed. The voter turnout was calculated by taking the total votes made for president in the 2012 election and dividing it by the Citizen Voting Age Population estimated by the US Census Bureau (2012).

The independent variable indicator to be used to test Hypothesis 4 will have to measure the affordability of transportation and housing. By using the proprietary measurement of the Housing and Transportation Index dataset, the H+T index, this should be the most effective means of measurement (Center for Neighborhood Technology 2017). The H+T index is an indicator which takes the annual transportation and housing costs of the median income household, as a percentage of the median income. This indicator is measured on a 0-100 scale with 100 being a combined housing and transportation cost equal to 100 percent of the median income. By measuring these costs as a percentage of the median income we are able to also control for variation of income that might make these costs look bigger than they actually are. The county level median income is a better measure than the national median income measure because in the partisan migration theoretical framework of Hypothesis 4 potential households in search of a county to relocate to would likely factor in how much housing and transportation would cost not in comparison to what the salary they are currently making but in comparison to what they will be likely to make in their new job at the destination of relocation. One issue with this indicator might be that the cost of transportation includes personal vehicle as well as public transportation costs. This would seem to be a huge error since this article is focused on public transportation and not personal vehicle transport. However, based on the theoretical framework presented earlier from Nall's 2015 results on how highway infrastructure caused Republican sorting due to the higher income required to maintain a personal vehicle it should be expected that areas with public transit access will have lower transportation and housing costs in turn (Glaeser et al 2008). These two independent variables just discussed will also be used in the models for testing the other hypotheses but in an additive control variable capacity only since there has not been a theoretical framework developed to justify putting them both in each models as interactions.[4]

3.4. Control Variables

The control variable will be broken down into three categories. The first set of variables will be those derived from the social capital theories and include those variable that are expected to effect social capital with changes in geospatial features. These variables include annual vehicle miles traveled per household, and how compact the average neighborhood is. Measured on a 0 to 10 scale neighborhood compactness would be expected to effect walkability and social interaction potential, and suggests levels of urbanization and population density (Farber & Li 2013). The second set of variables are the voter migration variables that would be expected to effect a partisan more than a non-partisan in their decisions to relocate. The variables include housing and transportation costs, and proximity to jobs (Mummolo & Nall 2017). The third grouping of variables represent demographic controls which are the most common explanation of the vote by class, origin, and partisan identity theories of voting. These variables include percentage of non-white residents, median household income, percent of population with at least a bachelor's degree, and percent of the county population below the national poverty line. The final variable that will

[4] Refer to Appendix B for further explanations of the independent and control variable's means, standard deviations, and min/max values

be used for control purposes is a simple categorical variable that will be used to measure state fixed effects.

3.5 Models for Hypothesis Testing

In order to test Hypotheses 1 and 2 the absolute value of the dependent variable, partisan clustering, will have to be used. Using the magnitude of the Dissimilarity value and the value of transportation access should test to see if there is evidence to support the hypothesis. The following equations represent the models that will be used to test the hypotheses:

H1: $$|\text{Partisan Clustering}_{2012i}| = \beta_0 + \beta_t \log(\text{Transit Access}_i+1) + \beta\text{controls} + \varepsilon_i \quad (5)$$

H2: $$|\text{Partisan Clustering}_{2012i}| = \beta_0 + \beta_t \log(\text{Transit Access}_i+1) + \beta_p \text{ Political Participation}_{2012i} \quad (6)$$
$$+ \beta_{tp} \log(\text{Transit Access}_i+1) * \text{Political Participation}_{2012i} + \beta\text{controls} + \varepsilon_i$$

Support for the null hypothesis would come in the form of there being weak support for an increase in transit ridership covarying with a decrease in partisan clustering, or evidence of an inverse relationship. The test for the interactive hypotheses, H2.1 and H2.2, is represented in equation 6. As was stated earlier the county's voter turnout in the 2012 election will be used to test the interaction.

The main difference between the tests for Hypotheses 1 & 2, and Hypotheses 3 & 4 is that the dependent variable switches from an absolute value signifying only magnitude in Hypothesis 1, to a value able to signify the direction of the partisan sorting in Hypothesis 3 (equation 7):

H3: $$\text{Partisan Clustering}_{2012i} = \beta_0 + \beta_t \log(\text{Transit Access}_i+1) + \beta\text{controls} + \varepsilon_i \quad (7)$$

H4: $$\text{Partisan Clustering}_{2012i} = \beta_0 + \beta_t \log(\text{Transit Access}_i+1) + \beta_c (\text{H+T Costs})_i \quad (8)$$
$$+ \beta_{tc} \log(\text{Transit Access}_i+1) * (\text{H+T Costs}) + \beta\text{controls} + \varepsilon_i$$

The equation that will be used to test the interactive hypotheses 4.1 and 4.2 is represented in equation 8. As mentioned before the indicator for housing and transit costs in the interaction test will be the H+T index.

IV. Empirical Results

4.1 Transit Access and Partisan Clustering

To test the first hypothesis of this research on how transit access effects the magnitude of partisan clustering a regression analysis has been used. Negative coefficients for the log of commutes by transit would have to be observed for the hypothesis to have sufficient evidence in its support.

VARIABLES	(Model 1) Absolute Value Dissimilarity	(Model 2) Absolute Value Dissimilarity	(Model 3) Absolute Value Dissimilarity	(Model 4) Absolute Value Dissimilarity
log(%transit commutes)	0.288	2.603**	0.581	8.531**
	(0.667)	(1.043)	(1.109)	(3.725)
Voter Turnout		-0.0979**	0.304***	0.410***
		(0.0401)	(0.0652)	(0.0807)
Vehicle miles traveled/household		0.00148***	0.000990***	0.00101***
		(0.000229)	(0.000255)	(0.000255)
Compact neighborhood score		1.948***	2.148***	2.151***
		(0.481)	(0.576)	(0.576)
Job Access score		0.197	0.455	0.511
		(0.369)	(0.413)	(0.413)
Housing+Transit costs %		0.270***	0.147**	0.140**
		(0.0608)	(0.0679)	(0.0679)
% Nonwhite voters			-0.102***	-0.113***
			(0.0376)	(0.0379)
County median income, % of Nat'l			0.0163	0.0202
			(0.0389)	(0.0389)
% below poverty line			0.481***	0.517***
			(0.154)	(0.154)
Education, % completed Bachelors			-0.227***	-0.217***
			(0.0676)	(0.0677)
log(% Transit Commutes)*Voter Turnout				-0.138**
				(0.0619)
State Fixed Effects:			YES	YES
Constant	23.22***	-29.11***	-33.20***	-40.37***
	(0.617)	(8.198)	(9.912)	(10.41)
Observations	1,801	1,801	1,801	1,801
R-squared	0.000	0.037	0.195	0.198

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.1.1 *Models for testing Hypotheses 1 and 2. The regression table with the bivariate model (1); multivariate additive models with: social interaction potential and relocation rational variables (2), and all variables (3); and the multivariate interactive model (4)*

As the results of Model 1 in table 4.1.1 suggest there is little support for transit access reducing partisan clustering, with the sign of the coefficient being positive rather than negative, having a low magnitude, and statistical insignificance. Model 1 shows that the effect of transit access on partisan clustering is statistically and substantively insignificant. Once the social capital and social interaction potential variables are controlled for (Model 2) the effects of transit access become more significant, but still in the opposite direction hypothesized. It should also be noted that the direction of the constant switches after the controls are added from positive to negative. In model 3 the coefficient for transportation access returns to insignificant levels with the sign of the coefficient remaining in the opposite direction suggested in Hypothesis 1.

The annual vehicle miles traveled was as expected a strong effect on clustering, with a meager 30 mile round trip commute in a single occupancy vehicle equating to an annual vehicle miles traveled of 7,500 and a subsequent expected effect on the absolute value of dissimilarity to increase clustering by 7.5 percentage points for models 3, 4, and 5. This was assumed through the social interaction potential theories. The density effect (Compact neighborhood score), like the effect of transit access, do not appear to have the effect theorized under social interaction potential. The demographic variables introduction into the models as well as the state fixed effects in models 3 and 4 have a substantial increase on explaining the variance observed as expected. The transit access effect on the absolute value of partisan clustering explains much less of the variance observed. However to better understand how transit access and political participation interact to produce and effect on partisan clustering a plot of the marginal effect will be needed and will test Hypothesis 2.

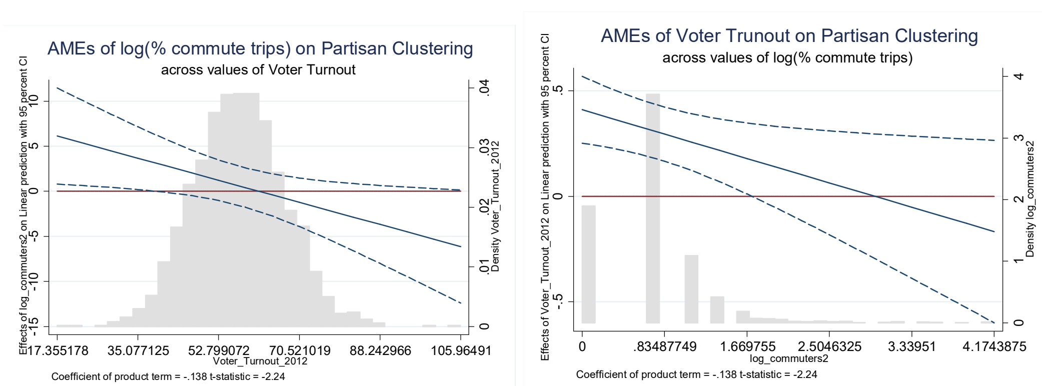

Figure 4.1.1 *The Average Marginal Effects plots of transit access on partisan clustering across values of voter turnout (H2.1; Left), and the marginal effect of voter turnout on partisan clustering across values of transit access(H2.2; Right). Both plots are calculated from the regression in model 4 of Table 4.1.1.*

The marginal effects of transit access on partisan clustering can be seen to be positive at the lowest values of voter turnout, and negative at the highest values of voter turnout. This is as was predicted in Hypotheis 2.1, however this interaction is shown to be insignificant as the confidence intervals never really move from the zero marginal effect line. The marginal effect of voter turnout on partisan clustering is also positive when transit access is low and negative when transit access is high. However the marginal effect of the voter turnout on partisan clustering is shown to only have a partisan clustering effect below the log of transit access value of 1.669. Above this value there is no statistical significance. The direction of the interaction testing Hypothesis 2.2 suggests that there is some support that with higher levels of transit access the marginal effect of political participation on partisan clustering is moderated. However the explicit prediction made in H2.2 that at higher levels of transit access the marginal effect of political participation on partisan clustering is negative is not shown here. The margins plot in Figure 4.1.1 shows instead that higher levels of transit access only weaken—but not reverse—the positive marginal effect political participation has on partisan clustering.

4.2 Transit access and Democratic Party Clustering

VARIABLES	(Model 5) Democratic Dissimilarity	(Model 6) Democratic Dissimilarity	(Model 7) Democratic Dissimilarity	(Model 8) Democratic Dissimilarity
log(%transit commutes)	17.14***	12.53***	0.144	-9.543**
	(0.965)	(1.499)	(1.082)	(4.562)
Housing+Transit costs %		0.0129	-0.300***	-0.450***
		(0.0874)	(0.0662)	(0.0952)
Voter Turnout		0.276***	0.331***	0.347***
		(0.0576)	(0.0636)	(0.0640)
Vehicle miles traveled/household		-0.00240***	-0.000226	-0.000349
		(0.000329)	(0.000249)	(0.000255)
Compact neighborhood score		0.0856	0.160	0.508
		(0.691)	(0.562)	(0.584)
Job Access score		-1.047**	0.576	0.419
		(0.531)	(0.403)	(0.409)
% Nonwhite voters			1.187***	1.190***
			(0.0367)	(0.0367)
County median income, % of Nat'l			-0.261***	-0.264***
			(0.0380)	(0.0379)
% below poverty line			0.574***	0.563***
			(0.150)	(0.150)
Education, % completed Bachelors			0.971***	0.959***
			(0.0660)	(0.0661)
log(%transit commutes)*H+T costs %				0.177**
				(0.0811)
State Fixed Effects?			YES	YES
Constant	-24.74***	26.23**	-44.77***	-33.70***
	(0.892)	(11.78)	(9.672)	(10.91)
Observations	1,801	1,801	1,801	1,801
R-squared	0.149	0.191	0.689	0.689

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.2.1 *Models for testing Hypotheses 3 and 4. The regression table with the bivariate model (5); multivariate additive models with: social interaction potential and relocation rational variables (6), and all variables (7); and the multivariate interactive model (8).*

The third hypothesis to be tested differs from the first hypothesis test in that for this test the directional variable for clustering will be used rather than the non-directional absolute value of the Dissimilarity value. The hypothesis assumes that as transit access increases so does Democratic party clustering. This will be observed through positive coefficients, while alternatively coefficients with negative signs suggest Republican clustering.

Models 5 and 6 in the above table show significant positive coefficients for the effect of transportation access on Democratic partisan clustering. This suggests that for every one unit increase in the log of the percent of commutes by transit that Democratic party clustering increases by more than 10 percentage points. Using the dissimilarity value this means that 5 percent of the population in the county would have to relocate to another county in order for the county to be said to have no clustering in comparison to the state's partisan makeup. In Model 7 after the introduction of the state fixed effects and demographic control variables—which are more common for explaining reasons for the vote—the effect of transit access becomes insignificant in comparison. However the sign on this insignificant coefficient retains the hypothesized direction and so could still be said to maintain some support for the second hypothesis. The interactive model (8) with all the control variables still shows a statistically and substantively significant coefficient for both independent variables and their interactive term. However the sign on the interactive model's transportation access coefficient switches from the additive models (6 and 7). To better understand the interactive relationship the margins plot in Figure 4.2.1 should be helpful.

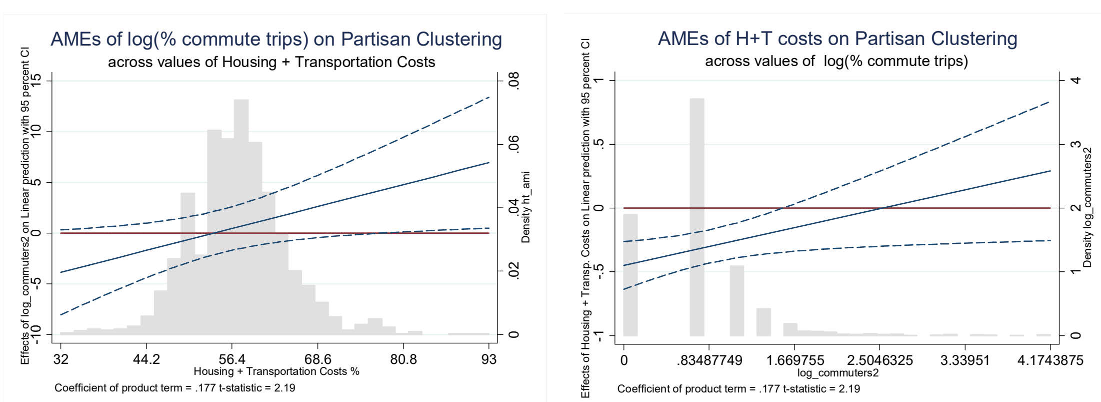

Figure 4.2.1 *The Average Marginal Effects plots of transit access on partisan clustering across values of housing and transportation costs (H4.1; Left), and the marginal effect of housing and transportation costs on partisan clustering across values of transit access (H4.2; Right). Both plots are calculated from the regression in model 8 of Table 4.2.1.*

The Average Marginal Effects plots in figure 4.2.1 show the additional effect that the interactive term has on partisan clustering. In the theoretical framework section of this paper Hypothesis 4.1 predicted that the marginal effect of transportation access on Democratic partisan clustering would be positive at low values of housing and transportation costs and negative at high values of housing and transportation costs. This effect is not reflected in the evidence presented here. Instead we observe the opposite, that at low values of housing and transportation costs there is a negative marginal effect by transportation access on Democratic partisan clustering and at high values of housing and transportation costs a positive effect. This would suggest that places with lower housing and transportation costs would experience partisan clustering in the Republican party direction when transportation access increases rather than in the democratic clustering direction, but the plot shows that nowhere in the interaction plotted is the relationship statistically significant. In the second margins plot showing the marginal effect of housing and transit costs on Democratic partisan clustering across values of transit access we can observe the predicted relationship from Hypothesis 4.2. Here the marginal effect of housing and transportation costs on Democratic partisan clustering is negative for the values below the value of 1.6 of the log of transit ridership. This effect is only observed to be positive when transportation access is at its highest but should not be trusted because the results above the 1.6 value of the log of transit access is not statistically significant. This suggests that with lower transit ridership levels the additional effect of housing and transportation costs will have the effect of contributing to Republican partisan clustering, while at higher levels of transit access there will be a weaker additional effect towards Republican partisan clustering by housing and transit costs.

The evidence presented in Figure 4.2.1 also goes against the relation between affordability, class theory and partisan clustering suggested by other researchers (Nall 2015; Glaesser et al 2008). An alternative understanding of this interaction would support the notion that transit access causes cost increases (Bannerjee et al 2012) and that class voting does not explain the vote in the US contrary to theorists that still argue in support of it (Evans 2000). However it could still be argued that greater housing and transportation costs as a percentage of income reflect not the upper and middle class but a lower class and thus the Hypothesis 4.1 was wrong to predict that the marginal effect of transit access on Democratic partisan clustering would be negative when costs for housing and transportation costs are highest. Overall the results presented here are too weak to support Hypothesis 4.1, yet provide some support for rejecting the null hypotheses of Hypotheses 3 and 4.2 .

4.3 Discussion

The analysis provided in the previous sections test the hypotheses presented in this research. The test for the first hypothesis found no support for the hypothesis that transit access decreases partisan clustering. Ideally further research would be needed to provide further support, but there is not enough support to reject the null hypothesis. The test of the Hypothesis 2.2 found some statistical and substantial support. The predictions made of the interaction returned not exactly as expected that the marginal effect of transit access on partisan clustering is negative when political participation is higher, but did find a decrease in the strength of the positive marginal effect political participation can have on partisan sorting with higher transit access values. This interaction highlights the social interaction and social capital building potential of transportation access.

The test for the third hypothesis as well found some support for the hypothesis that transit access increases democratic partisan clustering in the absence of certain controls, but that support weakened after controlling for more likely causes of Democratic partisan clustering. This test found tentative support to reject the null hypothesis.

The test of Hypothesis 4.1 did not find statistical nor substantive support, and did not return as expected that the marginal effect of transit access on partisan clustering would be negative when housing and transportation costs are higher. Part of this incorrect prediction could be from a misapplied theoretical framework, but also from improper operationalization of the affordability variable by using an indicator that included all transportation methods costs rather than just public transit costs.

These tests were then checked for robustness using the alternative indicators discussed earlier on. In summary, most of the results from these checks mirror the empirical results that were presented here. The one notable, although minor, exception was that the robustness check for the test of Hypothesis 1 suggested greater support for the hypothesized relationship but likely not enough to reverse the conclusions already made of that hypothesis test. The complete results, and some additional comments can be viewed in Appendix A.

V. Conclusion:

This study does not find support for the ability of public transit use to decrease partisan clustering, but rather finds some support for the hypothesis that an increase in transportation access correlates to an increase in Democratic partisan clustering. This result corroborates the findings of prior researchers on how the structural incentives for relocation cause partisan clustering, rather than the self-segregation of partisans (Mummolo & Nall 2017; Nall 2015). These results however would seem to give incentive for the Republican party to oppose public transit access and to draw district boundaries in a way to isolate this element. However, the effects observed on Democratic partisan clustering should not be overstated for their impact is minimal in comparison to traditional explanations for partisan identity formation and voting behavior.

Additionally, this study finds minor support for the claim made by social interaction theory that increases in space for social interaction cause political participation to have a tolerance inducing effect (Pettigrew 1998; Farber & Li 2013). This we observed in the H2.2 test showing the marginal effect of political participation on partisan spatial segregation to decrease within a range of transit access increases. Whether this effect was brought about primarily from social capital acting a bridge to foster intergroup contact in heterogenous communities, or whether it was a cause of homogenous communities increasing an insulated social capital through contact cannot be differentiated in this research (Currie & Stanley 2003; Putnam 2007). In actuality, the ability of transit riders to confront non-partisans and others of divergent backgrounds and views from their own is context dependent on the transit route having stops in and use by non-partisan communities. Further research should be done with smaller units of analysis to thoroughly measure the extent of this interaction between transit ridership as a space for potential intergroup social interaction and political participation.

How transit access effects partisan clustering is not a concept restricted to the USA. Further research should be done to see whether or not the effects of access to mass transit on partisan clustering is observed in a similar way outside the USA. Findings would likely reflect the role of the electoral system in their outcomes, with countries with single member winner take all districts likely under similar pressures of the redistricting process, and countries with multi-member districts and proportional representation likely experiencing less partisan clustering associated with transit access. An examination of partisan clustering in the rest of the world might bring further support for structural and electoral systems explanations for the observed phenomenon.

VI. Bibliography:

Abramowitz, A. (2010). The disappearing center: Engaged citizens, polarization, and American democracy. Yale University Press.

Banerjee, A., Duflo, E., & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China (No. w17897). *National Bureau of Economic Research*.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*.

Berry, W. D., Golder, M., & Milton, D. (2012). Improving tests of theories positing interaction. *The Journal of Politics*, 74(3), 653-671.

Bishop, B. (2009). The big sort: Why the clustering of like-minded America is tearing us apart. Houghton Mifflin Harcourt.

The Center for Neighborhood Technology (2016). The Housing and Transit Affordability Index: Download County Data by State. http://htaindex.cnt.org/download/data. Retrieved: 08 June 2017. Chicago, IL.

The Cook Political Report. (2017) 2018 House Race Ratings for February 2, 2017. http://cookpolitical.com/house/charts/race-ratings. Published 02/02/2017. Accessed 15/03/2017

Currie, G., & Stanley, J. (2008). Investigating links between social capital and public transport. *Transport Reviews*, *28*(4), 529-547.

Dyck, J. J., & Pearson-Merkowitz, S. (2014). To know you is not necessarily to love you: The partisan mediators of intergroup contact. *Political Behavior*, *36*(3), 553-580.

Evans, G. (2000). The continued significance of class voting. *Annual review of political science*, 3(1), 401-417.

Farber, S., & Li, X. (2013). Urban sprawl and social interaction potential: an empirical analysis of large metropolitan regions in the United States. *Journal of Transport Geography*, 31, 267-277.

Fainstein, S. S. (2005). Cities and diversity should we want it? Can we plan for it?. *Urban affairs review*, *41*(1), 3-19.

Federal Elections Commission (2013). Federal Elections 2012: Election Results for the U.S. President, the U.S. Senate and the U.S. House of Representatives. Federal Elections Commision: Washington. Retrieved on 16 June 2017. https://transition.fec.gov/pubrec/fe2012/federalelections2012.pdf

Feinberg, M., Tullett, A. M., Mensch, Z., Hart, W., & Gottlieb, S. (2017). The political reference point: How geography shapes political identity. *PLoS ONE*, 12(2), e0171497. http://doi.org/10.1371/journal.pone.0171497

Fiorina, M. P., & Abrams, S. J. (2012). Disconnect: The breakdown of representation in American politics (Vol. 11). University of Oklahoma Press.

Gimpel, J., Hui, I., Tam Cho, W. (2013). Voter Migration and Geographic Sorting of the American Electorate. *Annals of the Association of American Geographers*, 103 (4), 856-870

Glaeser, E. L., Kahn, M. E., & Rappaport, J. (2008). Why do the poor live in cities? The role of public transportation. *Journal of urban Economics*, 63(1), 1-24.

Hajnal, Z. L., & Lewis, P. G. (2003). Municipal institutions and voter turnout in local elections. *Urban Affairs Review*, 38(5), 645-668.

Huckfeldt, R., & Sprague, J. (1987). Networks in context: The social flow of political information. *American Political Science Review*, *81*(04), 1197-1216.

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual review of psychology*, *60*, 307-337.

Lang, C., & Pearson-Merkowitz, S. (2015). Partisan sorting in the United States, 1972-2012: New evidence from a dynamic analysis. *Political Geography*, 48, 119–129.

Manza, J., & Brooks, C. (1999). Social cleavages and political change: Voter alignments and US party coalitions. OUP Oxford.

Marschall, M., Shah, P., & Ruhil, A. (2011). The study of local elections. *PS: Political Science & Politics*, 44(01), 97-100. doi:10.1017/S1049096510001940

McGovern, T., & Larson, S. (2017). US president county-level election results for 2012 and 2016. https://github.com/tonmcg/County_Level_Election_Results_12-16. Retrieved: 05 May 2017. Download: https://github.com/tonmcg/County_Level_Election_Results_12-16/blob/master/US_County_Level_Presidential_Results_08-16.csv. Retrieved: 05 May 2017.

Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2012). How Ideological Migration Geographically Segregates and Polarizes Groups. *Journal of Experimental Social Psychology*, 51, 1–14

Mummolo, J., & Nall, C. (2017). Why Partisans Do Not Sort: The Constraints on Political Segregation. *The Journal of Politics*, 79(1), 000-000.

Nall, C. (2015). The Political Consequences of Spatial Policies: How Interstate Highways Facilitated Geographic Polarization. *The Journal of Politics*, 77(2), 394-406.

Office of the Clerk U.S House of Representatives, Karen L. Has (2017). Statistics of the Presidential and Congressional Election from official sources for the election of November 8, 2016. Retrieved from the web on: 06 June 2017. http://history.house.gov/Institution/Election-Statistics/Election-Statistics/. Washington: 2017.

Pettigrew, T. F. (1998). Intergroup contact theory. Annual review of psychology, 49(1), 65-85.

Putnam, R. D. (2001). Bowling alone: The collapse and revival of American community. Simon and Schuster.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture. *Scandinavian political studies*, *30*(2), 137-174.

Rodden, J. (2010). The geographic distribution of political preferences. *Annual Review of Political Science*, 13, 321-340.

Rogers, S. & Cage, F. (2012). "Full US 2012 election county-level results to download". 14 November 2012. https://www.theguardian.com/news/datablog/2012/nov/07/us-2012-election-county-results-download#data. Retrieved: 08 June 2017. The Guardian.

Sussell, J. (2013). New Support for the Big Sort Hypothesis: An Assessment of Partisan Geographic Sorting in California, 1992–2010. *Political Science & Politics*, 46 (4), 768-773.

Stolle, D., & Rochon, T. R. (1998). Are all associations alike? Member diversity, associational type, and the creation of social capital. *American Behavioral Scientist*, *42*(1), 47-65.

United States Census Bureau (2012). Small Area Income and Poverty Estimates: State and County Estimates for 2012. Retrieved on 14 June 2017, from: https://www.census.gov/did/www/saipe/data/statecounty/data/. Downloaded at: https://www.census.gov/did/www/saipe/downloads/estmod12/est12ALL.xls

United States Census Bureau (2009). Voting Age Population by Citizenship and Race (CVAP): CVAP from the 2005-2009 American Community Survey 5 year estimates. Retrieved on 12 June 2017, from: https://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html?eml=gd&utm_medium=email&utm_source=govdelivery.

United States Census Bureau (2012). Voting Age Population by Citizenship and Race (CVAP): CVAP from the 2008-2012 American Community Survey 5 year estimates. Retrieved on 12 June 2017, from: https://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html?eml=gd&utm_medium=email&utm_source=govdelivery.

United States Census Bureau (2017). Voting Age Population by Citizenship and Race (CVAP): CVAP from the 2011-2015 American Community Survey 5 year estimates. Retrieved on 12 June 2017, from: https://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html?eml=gd&utm_medium=email&utm_source=govdelivery. Downloaded at: https://www.census.gov/rdo/pdf/CVAP_CSV_Format_2011-2015.zip. Published 01 February 2017.

USDA (2017). Economic Research Service: Educational attainment for the U.S., States, and counties, 1970-2015. Retrieved on 12 June 2017, from: https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/.

Walker, K. E. (2013). Political Segregation of the Metropolis: Spatial Sorting by Partisan Voting in Metropolitan Minneapolis-St Paul. *City & Community*, *12*(1), 35-55

Wessel, T. (2009). Does diversity in urban space enhance intergroup contact and tolerance?. *Geografiska Annaler: Series B, Human Geography*, *91*(1), 5-17.

Wong, D. W. (2003). Spatial decomposition of segregation indices: a framework toward measuring segregation at multiple levels. *Geographical Analysis*, *35*(3), 179-194

VII. APPENDIX A: Robustness Checks

7.1 Robustness checks for the H1 and H2 tests:

Table R1:

VARIABLES	(R1) Absolute Value Dissimilarity	(R2) scalr_votediff_2012	(R3) scalr_votediff_2012	(R4) scalr_DvalueAvgnatl	(R5) scalr_DvalueAvgnatl
log(annual transit trips+1)	0.0732		-1.262***		-3.630***
	(0.416)		(0.458)		(0.487)
log(%transit commutes)		-1.119		-4.741***	
		(0.736)		(0.784)	
Constant	23.24***	28.08***	30.57***	34.53***	40.61***
	(1.155)	(0.680)	(1.272)	(0.725)	(1.351)
Observations	1,801	1,801	1,801	1,801	1,801
R-squared	0.000	0.001	0.004	0.020	0.030

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The above table shows results that provide support for Hypothesis 1, contrary to the results of the original hypothesis test (compare to Model 1 of table 4.1.1). The coefficients here on transit access are all significant for the dependent variables calculated using the 2012 vote differential (R2 and R3) and dissimilarity value of the average election results of 2008, 2012, and 2016, and are all the same sign hypothesized. This could suggest that holding for leadership effects should have been used for the primary hypothesis testing, but could also be reflecting the difference between calculating the dissimilarity value with state totals verses national totals. Model R1 reiterates the results from section 4.1 that in relation to the state there is either insignificant support for H1 or significant support against H1 for transit access to decrease partisan clustering. However Models R2-R4 suggest that there is significant support for transportation access to decrease partisan clustering. It should be noted that there is little difference between the regression using the log of annual transit trips of the median income household and the log of the percent of commutes using transit.

Moving on to the multivariate models in R6-R7 we see that these results of the simple bivariate regression are no longer observed. When controlling for all the variables the sign on the coefficient for transit access switches and no longer supports the first hypothesis. These additive multivariate models should be compared along side Model 3 of Table 4.1.1, and the interactive model models (R9-R11) should be compared to Model 4 of Table 4.1.1. In Models R9-R11 we see the same results observed in Table 4.1.1 The coefficients of the transit access and voter turnout variables, including the interactive coefficients, are all of a similar magnitude and direction. Do to this similarity graphing the margins plot would be expected to produce the same results and is thus not necessary. The results for H2 are shown to be robust.

Table R2:

VARIABLES	(R6) scalr_DvalueAvgnatl	(R7) scalr_Dvalue2012state	(R8) scalr_votediff_2012	(R9) scalr_DvalueAvgnatl	(R10) Absolute Value Dissimilarity	(R11) scalr_votediff_2012
log(%transit commutes)	5.789***	0.581	4.936***	11.10**	8.531**	11.31***
	(1.131)	(1.109)	(1.107)	(4.382)	(3.725)	(3.722)
avg_VoterTurnout08_16	-0.00411			0.0630		
	(0.0689)			(0.0873)		
Voter Turnout		0.304***	0.217***		0.410***	0.302***
		(0.0652)	(0.0651)		(0.0807)	(0.0806)
Vehicle miles traveled/household	8.16e-05	0.000990***	-6.18e-05	7.50e-05	0.00101***	-4.66e-05
	(0.000261)	(0.000255)	(0.000255)	(0.000261)	(0.000255)	(0.000255)
Compact neighborhood score	3.325***	2.148***	3.060***	3.316***	2.151***	3.063***
	(0.586)	(0.576)	(0.575)	(0.586)	(0.576)	(0.575)
Job Access score	-2.040***	0.455	-1.422***	-2.031***	0.511	-1.377***
	(0.420)	(0.413)	(0.412)	(0.420)	(0.413)	(0.413)
Housing+Transit costs %	0.0167	0.147**	0.00860	0.0132	0.140**	0.00271
	(0.0692)	(0.0679)	(0.0678)	(0.0692)	(0.0679)	(0.0678)
avg_nonwhiteCVAP	-0.422***			-0.429***		
	(0.0385)			(0.0389)		
County median income, % of Nat'l	0.0982**	0.0163	0.0725*	0.103**	0.0202	0.0756*
	(0.0397)	(0.0389)	(0.0389)	(0.0398)	(0.0389)	(0.0389)
% below poverty line	-0.0667	0.481***	0.113	-0.0435	0.517***	0.142
	(0.157)	(0.154)	(0.154)	(0.158)	(0.154)	(0.154)
Education, % completed Bachelors	-0.561***	-0.227***	-0.444***	-0.556***	-0.217***	-0.436***
	(0.0694)	(0.0676)	(0.0675)	(0.0695)	(0.0677)	(0.0676)
% Nonwhite voters		-0.102***	-0.268***		-0.113***	-0.277***
		(0.0376)	(0.0376)		(0.0379)	(0.0379)
log(annual transit trips+1)*Voter Turnout						
log(%transit commutes)*Voter Turnout					-0.138**	-0.111*
					(0.0619)	(0.0618)
log(annual transit trips+1)*avg_VoterTurnout08_16						
log(%transit commutes)*avg_VoterTurnout08_16				-0.0891		
				(0.0710)		
State Fixed Effects?	YES	YES	YES	YES	YES	YES
Constant	52.06***	-33.20***	28.89***	47.75***	-40.37***	23.14**
	(10.13)	(9.912)	(9.900)	(10.70)	(10.41)	(10.40)
Observations	1,801	1,801	1,801	1,801	1,801	1,801
R-squared	0.407	0.195	0.341	0.408	0.198	0.342

7.2 Robustness checks for the H3 and H4 tests:

Table R3:

VARIABLES	(R12) Dvalue_2012DRstate	(R13) vdifferentialDR_2012	(R14) vdifferentialDR_2012	(R15) Dvalue_avgDR_Natl	(R16) Dvalue_avgDR_Natl
log(%transit commutes)		24.46***		26.69***	
		(1.010)		(1.012)	
log(annual transit trips+1)	10.60***		14.83***		16.25***
	(0.603)		(0.635)		(0.637)
Constant	-40.12***	-33.46***	-54.65***	-41.26***	-64.54***
	(1.673)	(0.933)	(1.763)	(0.936)	(1.768)
Observations	1,801	1,801	1,801	1,801	1,801
R-squared	0.147	0.246	0.233	0.279	0.266

In the above table (to be compared with the results of Model 5 in Table 4.2.1) the coefficients of models R12-R16 are all shown to be significant, and in the direction hypothesized. There is shown to be little difference between the selection of the dependent or independent variable, so this simple bivariate test can be said to be robust.

In the Table below the additive Models R17-R19 should be compared to Model 7 of Table 4.2.1, and the interactive Models R20-R22 should be compared to Model 8 of Table 4.2.1. All of the models of important coefficients in the model are similar in magnitude and direction to the coefficients presented in the regression for the second hypothesis test. This suggests that the results reached in testing H3 and H4 are robust.

Table R4:

VARIABLES	(R17) Dvalue_avgDR_ Natl	(R18) Dvalue_2012DR state	(R19) vdifferentialDR_ 2012	(R20) Dvalue_avgDR_ Natl	(R21) Democrati c Dissimilar ity	(R22) vdifferentialDR_ 2012
log(%transit commutes)	-0.249	0.144	0.153	-13.24***	-9.543**	-10.50**
	(0.994)	(1.082)	(1.024)	(4.184)	(4.562)	(4.314)
Housing+Transit costs %	-0.318***	-0.300***	-0.282***	-0.518***	-0.450***	-0.446***
	(0.0608)	(0.0662)	(0.0627)	(0.0872)	(0.0952)	(0.0900)
Voter Turnout		0.331***	0.324***		0.347***	0.342***
		(0.0636)	(0.0602)		(0.0640)	(0.0605)
avg_VoterTurnout08_1 6	0.479***			0.500***		
	(0.0606)			(0.0608)		
Vehicle miles traveled/household	-0.000299	-0.000226	-0.000153	-0.000464**	-0.000349	-0.000288
	(0.000229)	(0.000249)	(0.000236)	(0.000234)	(0.000255)	(0.000241)
Compact neighborhood score	0.327	0.160	0.235	0.800	0.508	0.618
	(0.515)	(0.562)	(0.532)	(0.535)	(0.584)	(0.552)
Job Access score	0.791**	0.576	0.626	0.580	0.419	0.454
	(0.370)	(0.403)	(0.381)	(0.375)	(0.409)	(0.387)
avg_nonwhiteCVAP	1.183***			1.187***		
	(0.0338)			(0.0338)		
County median income, % of Nat'l	-0.255***	-0.261***	-0.250***	-0.259***	-0.264***	-0.253***
	(0.0349)	(0.0380)	(0.0359)	(0.0348)	(0.0379)	(0.0359)
% below poverty line	0.628***	0.574***	0.562***	0.611***	0.563***	0.549***
	(0.138)	(0.150)	(0.142)	(0.138)	(0.150)	(0.142)
Education, % completed Bachelors	1.111***	0.971***	0.918***	1.096***	0.959***	0.905***
	(0.0610)	(0.0660)	(0.0624)	(0.0610)	(0.0661)	(0.0625)
% Nonwhite voters		1.187***	1.139***		1.190***	1.142***
		(0.0367)	(0.0347)		(0.0367)	(0.0347)

	(1)	(2)	(3)	(4)	(5)	(6)
log(%transit commutes)*Housing+Transit costs %				0.238***	0.177**	0.195**
				(0.0743)	(0.0811)	(0.0767)
log(annual transit trips+1)*Housing+Transit costs %						
State Fixed Effects?	YES	YES	YES	YES	YES	YES
Constant	-85.29***	-44.77***	-67.71***	-70.45***	-33.70***	-55.55***
	(8.910)	(9.672)	(9.151)	(10.03)	(10.91)	(10.32)
Observations	1,801	1,801	1,801	1,801	1,801	1,801
R-squared	0.797	0.689	0.774	0.799	0.689	0.775

VIII. APPENDIX B: Data

Variable Name	Explanation	Observ-ations	Mean	Standard Dev.	Min	Max
Dependent Variable Indicators:						
Democratic Dissimilarity (Dvalue_2012DRstate)	the Dissimilarity value for the 2012 election. This is calculated using each party's percentage vote share in the county and their vote sharei in the state for the presidential election of 2012.	1801	-12.41772	25.79552	-77.16631	93.06157
Dvalue_2012DRNatl	the Dissimilarity value for the 2012 election. This is calculated using each party's percentage vote share in the county and their vote share for the national election 2012. In Excel the formula used was "=((1/2)*((BY2/51.06)-(CA2/47.2)))"	1801	-20.11289	29.2206	-93.17394	81.921
Dvalue_avgDR_Natl	the Dissimilarity value for the average of the 2008, 2012, and 2016 elections. This is calculated using each party's average percentage vote share in the county and their average vote share for the national elections in the three presidential elections observed. In Excel the formula used was "=((1/2)*((CQ2/50.723)-(CR2/46.313)))"	1801	-22.06895	29.38336	-149.2118	84.51061
vdifferentialDR_2012	The 2012 presidential vote differential (calculated by subtracting the Republican candidate's vote share by the Democratic candidate vote shar)	1801	-15.87005	28.66769	-87.52	84.24
AVGppointdifferential	An average of the 2008, 2012, and 2016 presidential elections vote differentials	1801	-16.94371	28.42697	-129.5033	86.29333
Absolute Value Dissimilarity (scalr_Dvalue2012state)	The absolute value of "Dvalue_2012DRstate"	1801	23.42728	16.44846	0.050446	93.06157
scalr_Dvalue2012Natl	The absolute value of "Dvalue_2012DRNatl"	1801	29.72973	19.34551	0.0179002	93.17393
scalr_DvalueAvgnatl	The absolute value of " Dvalue_avgDR_Natl "	1801	31.1223	19.53343	0.0066109	149.2118
scalr_votediff_2012	The absolute value of " vdifferentialDR_2012"	1801	27.27515	18.15181	0.03	87.52
scalr_votediff_avg	The absolute value of "AVGppointdifferential"	1801	27.83284	17.89536	0.0433333	129.5033
Independent Variable Indicators:						
pct_transit_commuters_ami	Transit Ridership % of Workers for the Regional Typical Household	1801	1.620766	3.72111	0	64
transit_trips_ami	Annual number of Transit Trips for the regional typical household	1801	22.59189	51.525	0	893
log_transittrips2	log(x+1) transformation for right skew correction, for "transit_trips_ami"	1801	2.614565	0.932181 4	0	6.795706
log(%transit commutes) (log_commuters2)	log(x+1) transformation for right skew correction for "pct_transit_commuters_ami"	1801	0.718864 9	0.581177	0	4.174387
logcommuters	log transformation of % of commutes by public transit	1356	0.396981 1	0.654038	0	4.158883

logtransittrips	log transformation of annual number of transit trips for the regional typical household	1784	2.524008	1.002799	0	6.794587
Interaction variables:						
Voter Turnout	The percentage of the Citizen Voting Age Population that turned out to vote in the 2012 presidential election based on estimates from the Census Bureau	1801	57.30824	9.99002	17.35518	105.9649
avg_VoterTurnout08_16	County average CVAP voter turnout for the presidential elections of 2008, 2012, and 2016 (percentages in over 100 could have been caused by the rare underestimating of population growth or retention for a county)	1801	59.07124	9.364595	16.24518	119.1801
Housing+Transit costs % (ht_ami)	Housing + Transportation (personal and public transit) Costs as a % of Income for the Regional Typical Household	1801	57.17601	6.921031	32	93
Relocation/Geospatial Controls:						
Job Access score	Job Access Score (0-10), based on access to jobs and variety of employment types.	1801	3.323876	1.841318	0	9.9
Compact neighborhood score	Compact Neighborhood Score (0-10). A combined measure of household density and walkability to transit and services	1801	1.334703	1.183226	0	9
Vehicle miles traveled/household	Annual vehicle miles traveled per regional typical household; a measure of transit access; an increase should decrease voter turnout; should increase Republican clustering (Nall 2015)	1801	25322.73	2816.729	1163	36396
Demographic Controls:						
% Nonwhite voters	The percentage of a county's population that does not identify as "White only", based on Census Bureau data	1801	18.80839	17.50829	0	97.54143
County median income, % of Nat'l	The county's median household income as a percent of the national median household income. Numbers below 100 are below the national median income.	1801	93.41756	24.11745	45.46923	236.0281
% below poverty line	The percent of the county's population below the federal poverty line.	1801	16.3628	6.061142	3.1	45
avg_nonwhiteCVAP	percent of county's population that doesn't identify as "White only"; averaged from 2008, 2012, and 2016 Census data	1801	18.78255	17.39108	0.6633876	97.87195
Education, % completed Bachelors	The percent of the population having completed a Bachelors degree or higher. Taken from USDA 2017.	1801	23.0286	9.963146	6.4	78.8
State fixed effects, Dummy Variable:						
State Fixed Effects	The unique code (State FIPS) that identifies the state in which a county is located. Used for creating a dummy variable for state fixed effects.	1801	30.48251	15.13542	1	56
Unit of Analysis Descriptors:						
county_FIPS	A five digit code which identifies the state and the county, the first two digits are the state's FIPS code (S) while the following three digits represent the county (C) code (ex: SS-CCC)	1801	30584.96	15158.41	1001	56041
population	Resident population in the county in 2013	1801	162037.9	406611.5	382	9888601
households	number of households in county in 2013	1801	59941.17	142430.4	159	3228672
land_acres	Area of county in acres	1801	579577.2	885973.9	1271.3	1.28E+07

The primary dataset used was merged into a single excel sheet through the use Microsoft Excel's VLOOKUP function. The transportation portion of the data was downloaded individually through the Housing+Transportation Affordability Index website in

.csv format, and each state's dataset was merged into a single .csv sheet through Window's "Command Prompt" command:

```
C:\..., copy *.csv merged.csv
```

This data comes from the Center for Neighborhood Technology's 2013 Housing and Transportation Affordability Index (2016). Added to this dataset was the vote totals for each county in the 2008, 2012, and 2016 presidential elections. This data was aggregated by a third party source which reports that its original source was The Guardian and Townhall.com (McGovern & Larson 2017). A brief check on their sources uncovered their Guardian source to be viable for the main year under observation for this research, the 2012 presidential election, but the existence of the original Townhall.com data could not be ascertained (The Guardian 2012). For future research knowledge of Geographic Information Systems software such as ARCGIS would be extremely useful for aggregating official data sources with geographic components. The remaining control variable data came from the United States Census Bureau's Citizen Voting Age Population estimates (2009; 20012; 2017). Data added to the primary dataset from this source included the raw data for calculating the voter turnout and the nonwhite population and for each county and for the election year. Demographic data on poverty levels and median household income comes from the Census Bureau's Small Area Income and Poverty Estimates (2012). Data on education levels for each county came from the USDA's Economic Research Service (2017). The statewide vote data for the 2012 presidential election was taken from the Federal Election Commission's official calculations and was used to calculate the state based 2012 Dissimilarity value (2013).

Calculating the absolute values of the dependent variables and the log of the independent variables were done through Stata commands. The remaining calculations that were necessary were done through Excel. These include the voter turnout, the average vote differential, the Dissimilarity values, and the percentage of the population that does not identify as white-only, to name a few. For compatibility reasons all of the variables measured as a percent in fraction form were multiplied by a hundred so that 1 means 1% throughout the dataset to avoid 1 meaning 100% for some of the data. This means, for example, that on all of the regression table results for all of the variables measured as percentages a coefficient of 5 would mean that a 1 percent increase in that variable would have a 5 unit increase, or 5 percent increase, on the dependent variable.