

# **A Direct-Observation Approach to Identify Small-Area Variation in Political Behavior: The Case of Income, Partisanship, and Geography**

Eitan Hersh  
Clayton Nall<sup>1</sup>

First Draft: August 24, 2012

This Draft: September 9, 2013

<sup>1</sup>Eitan Hersh is Assistant Professor, Department of Political Science, Yale University. Address: Institute for Social and Policy Studies, Yale University, 77 Prospect Street, P.O. Box 208209, New Haven, CT 06520-8209; Email: eitan.hersh@yale.edu; Phone: 203-436-9061. Clayton Nall is Assistant Professor, Department of Political Science, Stanford University. Address: 616 Serra Street Rm 100, Stanford CA 94305; Email: nall@stanford.edu; Phone: (650)725-4076. We thank David Broockman, Anthony Fowler, Andrew Gelman, Justin Grimmer, Lauren Davenport, Jacob Hacker, David Laitin, and participants in the Stanford Political Science Methods Workshop and University of Michigan Spatial Analysis in Social Sciences seminar for comments on previous drafts. Thanks to Yale's Institution for Social and Policy Studies and the Center for the Study of American Politics for financial support and to Steve Ansolabehere and Jonathan Rodden for sharing the Harvard Election Data Archive precinct data.

## **Abstract**

Political behavioral research on geographic variation typically employs national surveys and rarely digs below the congressional district level. The data used in such analyses are not suitable for detecting geographic patterns of shared political behavior, or “communities of interest,” in geographic data. Using large databases containing micro-level voter data, we identify the communities in which different types of voter behavior cluster geographically, without relying on the assumptions associated with survey research. As a motivating example, we examine variation in income-based voting across and within states. Using block-group-level party registration data and precinct-level election returns, we employ a combination of nonparametric and multi-level models to demonstrate that much state-by-state variation in income-based voting is driven by differences in geographic clusters that rarely encompass states and often cross state boundaries.

# 1 Introduction

Given the size and diversity of the United States, there has always been concern over the extent and consequences of geographic factionalism. How should we analyze geographic factionalism? Measuring American geographic fractionalization calls for geographically precise data, yet the postwar behavioral revolution in political science has led to the adoption of sampling techniques in which respondents' immediate geographic context is typically assumed to be unimportant (Blumer, 1948). Surveys are rarely designed to be representative at the substate level, despite the many politically important social and economic divisions that occur there. Americans are segregated by income (Reardon and Bischoff, 2011), race (Farley et al., 1994; Ellen, 2000), and partisanship (Glaeser and Ward, 2006; Cho, Gimpel and Hui, N.d.), and they often form easily identifiable geographic communities with political cultures based on their religious, economic, and racial composition (Gimpel and Schuknecht, 2004). Such regions often bear little resemblance to the state and national boundaries in which sample data are intended to be representative. Indeed, before adoption of survey-based measurement of political behavior, study of these substate communities was the default mode of studying political change. As a result, political scientists paid more attention to describing political communities than inferring individual-level behavior and attitudes.

In this paper, we show that the recent development of national voter files with accompanying block-group demographic information and geographically coded precinct data, permits political scientists to study geographic communities of interest at a low geographic level using simple methods and without resorting to assumptions that usually arise when attempting to make inferences about public opinion within small areas. With full censuses of voters at the block-group and precinct levels, we show that state boundaries poorly represent geographic variation in behavior and voter responses to economic interests. Using two large data sets and spatial econometric methods that have previously been applied to county-level data (Gimpel and Schuknecht, 2002*b*), we demonstrate that communities of interest as captured by political behavior measured in voter records bear little resemblance to state boundaries, and often appear well below county lines. We are able to estimate variation in political behavior and partisan

affiliation at a much lower level of aggregation, and with fewer modeling assumptions, than previous data have permitted.

We show that the modifiable areal unit problem (MAUP), a well-known problem in geography, is a substantively important problem for the study of sub-national politics. Relying on state boundaries as containers for political behavior makes findings vulnerable to how states are drawn, an instance of the modifiable areal unit problem. For some research questions, this is not a critical problem: if our purpose is to study state-level opinion on state-level politics (such as opinion on state-level policies like gay marriage (Lax and Phillips, 2009) or state-level elections), we must rely on survey data collected within state (or lower) boundaries. However, if we are interested in how common geographic behaviors cross state lines, there are few alternatives to data that provide continuous, or near continuous, geographic coverage.

As an example of this issue's importance we use these micro-level aggregate data to reevaluate state-level analyses of income-based voting. We show that many cross-state differences in income-based (or "class-based") voting can be explained by behavioral differences in a few easily identifiable regions unrelated to state lines. We adopt this motivating example because various forms of "false consciousness" of refusing to recognize one's "correct" class politics has been a topic of interest at least since Marx and to political scientists for at least the last decade (Frank, 2004; Bartels, 2006; Gelman et al., 2007, 2008). Scholars have noted that high-income white voters in some states are more likely to be Democrats, while low-income white voters in some states are more likely to be Republicans. Other scholarship has focused on the role of state-level racial context in inducing policy conservatism (e.g. Alesina and Glaeser 2004).

We show that geographic clustering of income-based partisan voting rarely follows state lines, and that there are two major clusters in which the income-voting relationship is much different than would otherwise be expected: in the Black Belt and other areas with rural minority poverty, and in the Mormon Corridor and other middle-income, religious areas across the Mountain West. These clusters do not neatly follow state lines. More fine-grained geographic data allow us to detect these areas with much more confidence than with nationally representative survey samples or with other research techniques based on individual-level data.

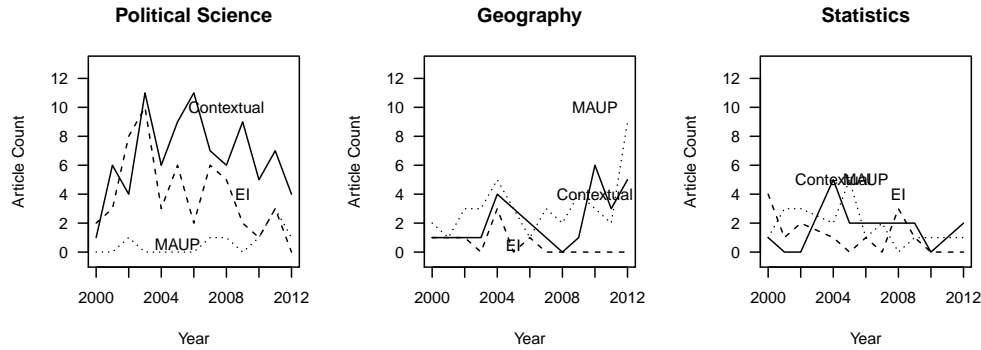
## **2 The Modifiable Areal Unit Problem and the Value of “Big Data” in the Study of Geographic Variation**

Political research on contextual effects and geographic behavioral variation is vulnerable to the Modifiable Areal Unit Problem (MAUP) (Openshaw and Taylor, 1981; Fotheringham and Wong, 1991): parameters estimated within geographic areas are sensitive to the choice of containers used in aggregation. Changing geographic scale (how large geographic units are) and zoning (where lines are drawn) can dramatically change estimated parameters. Commonly used containers, such as states and congressional districts, are often chosen by political scientists without regard to how the MAUP may affect inferences about geographic variation. While the MAUP has often been discussed as a problem only for political geography (King, 1997), the problem extends to all political scientists who aim to explain geographic variation in political behavior within geographic areas, whether they seek to estimate contextual effects or describe correlations within those areas. Informed by the MAUP, recent scholarship has demonstrated that the geographies of convenience (Census and political boundaries) used in contextual effects research rarely reflect individuals' perceived neighborhoods and communities (Wong et al., 2012). Similarly, extensive and conflicting scholarship on racial contextual effects has been shown to be sensitive to the choice of contextual geography (Cho and Baer, N.d.). Other research has shown that naturally formed communities, ranging from the “T-communities” formed by minor street networks (Grannis, 2009) to the mountainous stateless enclaves populated by communities of “hill people” rarely hew to state boundaries (Scott, 2010). The MAUP is not a mere statistical nuisance but a fundamental feature of political geography obscured in research that limits scholarship to boundaries adopted by the state (Scott, 1998).

Political scientists have tended not to be concerned with the MAUP, even after discussion of the issue in various articles and books on ecological inference in the late 1990s (King, 1997). As Figure 1 shows, compared to researchers in other fields that focus on geographic data, political scientists over the last ten years have barely discussed the issue. In comparison, political scientists have given considerable attention to contextual effects and ecological inference. Scholars in other fields, notably in geography,

have steadily given it more attention.

Figure 1: Inattention to the Modifiable Areal Unit Problem in Political Science



Number of articles on Google Scholar discussing “contextual effects,” “ecological inference,” and “modifiable areal unit problem” in three fields that work with geographic data. Journals selected from each discipline appear in Appendix Section A

Political scientists often accept given political boundaries as the bins in which they should analyze covariance. This is appropriate when substantive reasons exist to study particular geographic units. For example, public opinion within states explains the origins of state policies and statewide elections. But, if one’s purpose is to identify common contextual predictors (if not causes) of voter behavior, more justification is needed to accept states as boundaries of interest. State-level opinion is the result of data-generating processes arising out of different geographic clusters that vary within states (and usually crossing state lines). Even researchers studying political behavior over large areas should therefore be concerned with the data-generating process in subareas.

The modifiable areal unit problem becomes especially relevant when political scientists aim to use individual-level survey data to estimate effects within small geographic areas, an instance of the more general problem of small-area estimation (Rao, 2003; Park, Gelman and Bafumi, 2004). National surveys, and the methods used to analyze them, are highly constrained in estimating opinion within small areas. Canonical surveys such as the American National Election Studies and the National Annenberg Election Studies are designed to create nationally representative samples, not local ones. Methods

such as multi-level regression and poststratification (MRP) can be used to improve estimation (versus geographic disaggregation of survey data, and relative to relying on election-based proxies), but such methods rely heavily on non-survey data in small-area analysis. With the opt-in design used in many contemporary surveys, the data in some areas of the country are even sparser. In the Southern black belt, for example, an area of the country of geographic and theoretical distinction, most counties have fewer than 10 respondents in the 30,000+ respondent CCES.

A recent response to concerns over MAUP is to elicit respondents' self-identified neighborhood, producing the geographic "reality with which people engage, regardless of the accuracy or distortion of those perceptions" (Wong et al., 2012). This approach is a major advance in resolving what people mean when asked about their "neighborhood" or "community." While self-identified neighborhoods may be vectors for important social stimuli, however, this approach leaves several issues unsolved. Factors associated with objective, predefined contexts (e.g., unemployment rates in a metropolitan labor market area) may directly influence attitudes (Wong et al., 2012). Moreover, the direct elicitation method is, by design, suited to egocentric, individual-level analysis but are not suited to classifying behavior across geographic space.<sup>1</sup>

A second response is that MAUP is a theoretical problem with no empirically derived solution achieved through access to better data (King, 1997, 251–252). Another way to frame this critique is that no unsupervised method exists for discovering the "right" geographic boundaries to be used in analysis, a problem that is not limited to geography. In short, there is no single "right" context in which to study geographic variation, but there are answers that more or less effectively express the underlying geographic communities and clusters of interest. One option is to use supervised but data-driven techniques to reveal geographic patterns that do not follow state lines. This has been a widespread practice in political geography (e.g., Gimpel and Schuknecht (2004); Morrill, Knopp and Brown (2007, 2011)). The most common of these approaches is to use methods based on spatial autocorrelation, including Local Moran's  $I$ , which is used to identify spatial clusters of high and low outcome values.

---

<sup>1</sup>Such methods could be used to ascertain whether individuals who share geographic space also share perceptions of their neighborhood context, but this is a different question.

### 3 Motivating Example: Spatial Variation in Income-Based Voting

To demonstrate the significance of the MAUP for the study of geographic variation in voter behavior and to demonstrate the value of newly-available micro-scale aggregates, we focus on geographic variation in the correlation between neighborhood-level income and partisanship. We examine block groups, the lowest-level geography at which income is reported by the Census Bureau, and where we can link to precincts and registration records in which voting behavior and party registration behavior are reported by election authorities. We examine how the relationship between block-group/precinct partisanship and block-group/precinct income varies across state legislative districts, counties and states. We show that a set of distinctive sub-state regions explain cross-state differences in voter behavior.

Our approach differs from recent scholarship that has analyzed sparse survey data using multi-level regression and post-stratification (MRP). Such studies have tended to focus on *state-level* geographic variation (Gelman and Little, 1997; Park, Gelman and Bafumi, 2004; Lax and Phillips, 2009; Vigdor, 2006; Ghitza and Gelman, 2012; Lax and Phillips, 2009), since they depend upon use of Census micro-data to obtain proportions for poststratification, and such frequencies are rarely available at low levels. Analyses that aim to infer opinion below that level must borrow substantially from outside aggregate-level data (e.g., Warshaw and Rodden 2012; Ghitza and Gelman 2012; Tausanovitch and Warshaw 2011). As with other ecological inference strategies, if an area has no individual-level survey responses, an MRP technique will assign a predicted value based only on aggregate data and on the opinions of individuals who live elsewhere, making the technique highly dependent on an assumption of geographic independence. Rather than using such methods to improve estimation of opinion, we adopt a different inferential target (voters aggregated into block groups and precincts), and a behavioral rather than opinion-based outcome: voting and party registration. In doing so, we trump data sparsity issues, using a full census of voters and precincts from the Census Bureau, voter registration systems, and precinct-level election results. Such data are not as well suited as surveys for the study of public opinion, policy attitudes, or ideology, but are geographically extensive, allow detection of geographic behavioral clustering, and sidestep most small-area estimation problems associated with survey data.



We study variation of block groups and precincts within state house districts. We do so for several reasons. State house districts are the lowest level of representation that elect politicians responsible for major policymaking, most notably redistributive policymaking, which is rarely a function of city- or county-level governments (Peterson, 1981). They are especially useful in the study of metropolitan areas, which are generally contained within just a few populous counties (especially in the West) but are almost always divided into numerous legislative districts. In the post-*Baker v. Carr* era, these district lines rarely respect county lines. Finally, while they are, like state boundaries, political lines, they are typically as small as counties and especially in metropolitan areas permit more geographic precision than one could obtain using either states or counties.

Using substate data, we adopt multiple methods to show that deviations from income-based partisanship and voting are rooted in substate enclaves. First, plotting the relationship between district income and district partisanship and then block-group income and block-group partisanship, we learn that these measures co-vary not based on the wealth of the area they are in (as has been found regarding individual-level income and partisanship data within states), but rather they vary depending on the racial composition of the area. The income-party relationship is typically weak in predominantly white districts, regardless of state-level or district-level income context. Cross-state differences appear only in the portions of states composed by racially diverse districts. Finally, having shown that particular classes of districts within states engage in more or less income-based voting, we use local indicators of spatial autocorrelation to identify the geographic location of clusters, or the “spatial regimes,” within which income-based voting is much stronger (or weaker) than expected (Gimpel and Schuknecht, 2004).

### **3.1 Hypotheses**

Based on prior expectations from scholarship using county-level aggregate data, we hypothesize that differences in the block-group/precinct income-partisanship relationship (which we will short-hand as “income-based voting”) are not explained by demographics or culture alone, but are rooted in geographically specific areas in which voters belong to geographic communities that deviate substantially from typical income-based voting patterns.

We expect to see clustering of stronger income-based voting in racially diverse areas, especially where minorities are low-income. In such places, low-income voters are almost all Democratic (Dawson, 1995). In racially diverse areas where wealthier voters living near blacks and other racial and ethnic minorities are politically liberal, the relationship between income and partisanship will be weak. Where wealthy voters living near blacks are politically conservative, the relationship between income and partisanship will be quite strong. The former condition is likely to be especially common in urban areas with cosmopolitan voters (Gelman et al., 2008), but is much less likely outside of metropolitan areas, even in the so-called “rich/blue” states. The latter condition is prevalent in the South and less prevalent elsewhere. Consistent with the queries, “What’s the Matter with Kansas” and “What’s the Matter with Connecticut?” we expect a weaker party relationship in districts where both rich and poor whites have high levels of religiosity and in places where rich liberals share districts with poor racial minorities (Gelman et al., 2008; Frank, 2004).

### **3.2 Data**

We use two large data sets based on voter files and precinct-level election results to estimate the geographic extent of income-based voting. These data sources are summarized in Table 1. The first is a commercial database of 73 million voter records assembled by Catalist, a major political data vendor, providing data from the 29 states that require voters to register with a party or as independent/unaffiliated voters (Ansolabehere and Hersh, 2012). These data are published along with block-group-level and some individual-level characteristics for each voter. The second data set is 2008 presidential election returns from 185,000 precincts from 49 states that publish their precinct-level returns. This data allows us to estimate the relationship between precinct-level income and the 2008 two-party presidential vote (Ansolabehere and Rodden, 2011).

In the party registration dataset, Catalist represents block-group income as a binned categorical variable partitioned at \$20,000 intervals, ranging from \$20,000 or less to the Census top-code of \$200,000 and above. Each individual voter in the Catalist database is linked to the income bin of his or her block group. In this dataset, we can estimate the party registration distribution for all voters, for example, who

Table 1: Summary of Data Sources

	<b>Catalist Data</b>	<b>HEDA Data</b>
<b>Individual</b>	73,170,970 D. and R. Registrants	185,002 precincts
<b>Units</b>	with block group (BG) income measure	with BG income measure
<b>Aggregate</b>	State House Districts	Counties and State House Districts
<b>Units</b>	with aggregated BG inc. and race	with aggregated BG inc. and race
<b>Coverage</b>	29 Party Registration States	49 States with Published Data

live in block groups with median income less than \$20,000 in a particular state house district. Similarly, each voter is indicated by her block-group's racial composition. We represent district-level income by taking the unweighted mean of voters' block-group income, and the district percentage black by averaging the black population percentage of registered voters' block groups. We draw attention to the racial diversity of districts because, as we will see, the relationship between block-group income and block-group partisanship varies substantially with district-level race measures.

We also examine the income-party relationship within districts using precinct-level data from the Harvard Election Data Archive (hereafter, HEDA) (Ansolabehere and Rodden, 2011). To calculate precinct-level income and racial composition, race and population data were obtained from the ESRI block-group layer for the year 2000 (ESRI, 2008). Block-group median household income was obtained from the National Historical GIS (Fitch and Ruggles, 2003). To link these data to precincts, block groups were converted to points, and a weighted average of the points in each precinct was taken to generate the precinct-level demographic data. Precincts' state house district and county were coded by associating each precinct's centroid with the district or county polygon in which it fell (United States. Bureau of the Census, 2006).<sup>2</sup> This process yielded 185,002 precincts distributed across districts and counties in the 49 states. Summary statistics for these data appear in Appendix Table A-3.

<sup>2</sup>Centroids were calculated using the "Feature to Point" function in ArcGIS.

### 3.3 The Substate Origins of Income-Based Voting

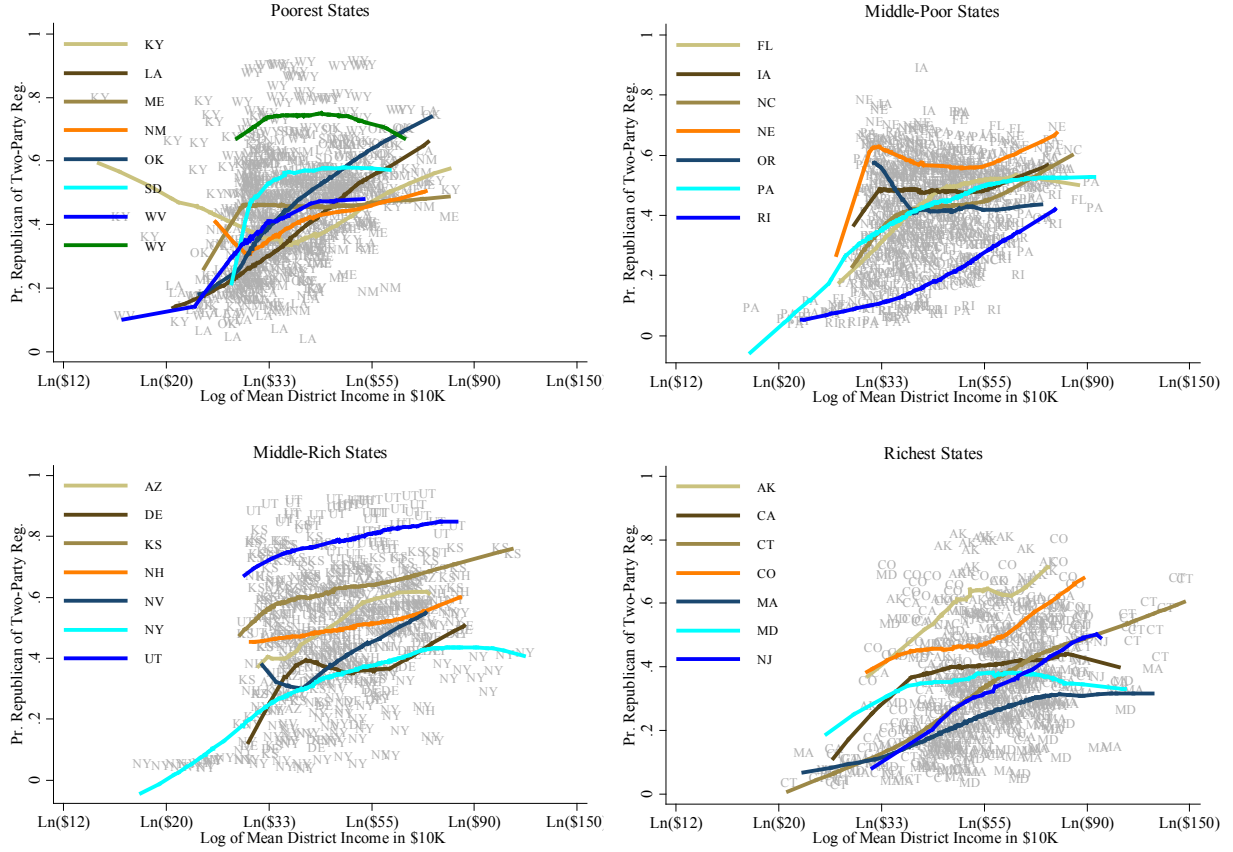
To demonstrate the importance of substate geographic heterogeneity and to summarize the Catalist data visually, we display the Republican share two-party registration as a function of district-level income in the 29 party-registration states. We graph this share as a function of average district income (Figure 2). Emulating Gelman et al. (2008), we partition party-registration states into four approximately evenly divided groups based on state income.<sup>3</sup> We plot the Republican proportion of two-party registrants against the logged mean district income, and then fit a lowess curve to each state.

Figure 2 demonstrates that important aspects of sub-state political geography are obscured in adopting state boundaries in the study of geographic variation. The first such aspect is that greater variation exists in partisan registration across districts within states than across states. Within states, the range of state-house-district incomes is as high as \$130,000. By contrast, the difference in average income between the richest and poorest state is only \$35,000. Second, if we assume that state contextual income is driving income-based voting, we might then expect the district income-party relationship to vary substantially across states. Despite cross-state differences in average district income, in 27 of the 29 party-registration states, richer districts are more Republican districts, regardless of state-level income. Third, the states with a combination of high-income Republican districts and very poor Democratic districts tend to have a mix of rich white and high-poverty minority (typically African-American, but also low-income Hispanic/Latino and American Indian) areas. These include rich (blue) states like New Jersey and Connecticut as well as poor (red) states like Louisiana and Oklahoma. States with a flat relationship—Wyoming, Kentucky, Iowa, Oregon, Utah and New Hampshire—are all less than 2% African-American. Even the most racially diverse of these states, Kentucky, had the South's lowest African-American population share (7.7%) in 2010 and few substantially segregated black areas. In states with poor Democratic districts but an otherwise flat income-party relationship—South Dakota, Nebraska, Delaware, and California—poor racial-minority districts exist but are only a small fraction of the state.

---

<sup>3</sup>State-level income is calculated by aggregating the block group median income value of the voters.

Figure 2: Regardless of State Income, Rich Districts are Republican Districts



Note: Figures are derived from the full population of registered Democrats and Republicans in party registration states. N=73,170,970. Average median household income is the unweighted average of the median household income of block groups in the district; partisan registration is obtained from individual-level voter file records.

We now turn to assessing income-based voting within districts as a function of block-group-level income and racial composition. We stratify state house districts based on their average income levels and their racial composition. In this analysis, we adopt minimal modeling assumptions, calculating the difference in partisan registration across income groups in different geographic settings using the full voter census by creating what are effectively visual representations of crosstabs. The large number of observations in the voter file allows us to overwhelm the small-area estimation problem by brute force

and plotting conditional means.

Figure 3 demonstrates the significance of district-level racial heterogeneity in explaining cross-state differences in income-based voting. The relationship between block-group income (in \$20,000 income intervals starting at \$20,000 and below and capped at \$100,000+) and the Republican proportion of two-party registrants within each income category is displayed.<sup>4</sup> These proportions are calculated and plotted for registered voters in block groups that fall within districts in three income categories: those with average block-group median income under \$40,000, \$40,000 to \$60,000, and more than \$60,000, each represented as a different line style. To demonstrate how variance with district-level context varies with state income, red lines represent voters in the wealthier 14 party-registration and purple lines the 15 poorer party-registration states. To test the racial composition intuition in Figure 2, each graph renders district-level racial composition category: 0 to 5 percent black, 5 to 10 percent black, 10 to 25 percent black, and 25 percent or more black. The main explanatory variable, block-group median household income, appears on the horizontal axis. This graph permits a reading of partisan affiliation in any cell defined by block-group income, district-level income and racial composition, and state-level income. For example, the solid red line in the upper left plot shows the relationship between block-group income and partisanship in districts that are poor, less than 5% black, and located in the rich states. In the Online Appendix (Table A-1), we conduct a version of this analysis using linear regression, which confirms the results presented here.

Variation in the income-party relationship within states is explained almost entirely by cross-state differences within racially diverse areas. In rich and poor, red and blue states, the slope of the income-party relationship in homogeneously white districts is remarkably similar, regardless of district-level income. This geographic finding would never be discovered by individual-level analyses within states. If state-level income explained voting behavior, we would expect the purple lines (poor states) to be steeper than the red lines (rich states). Likewise, if voters in poor districts were more likely to vote with their income interests, the solid lines would be steeper than the dotted lines. None of these relationships

---

<sup>4</sup>These bins were defined by Catalist. Because so few block groups have median household income above \$100,000, voters in block groups above this level were collapsed into the top category.

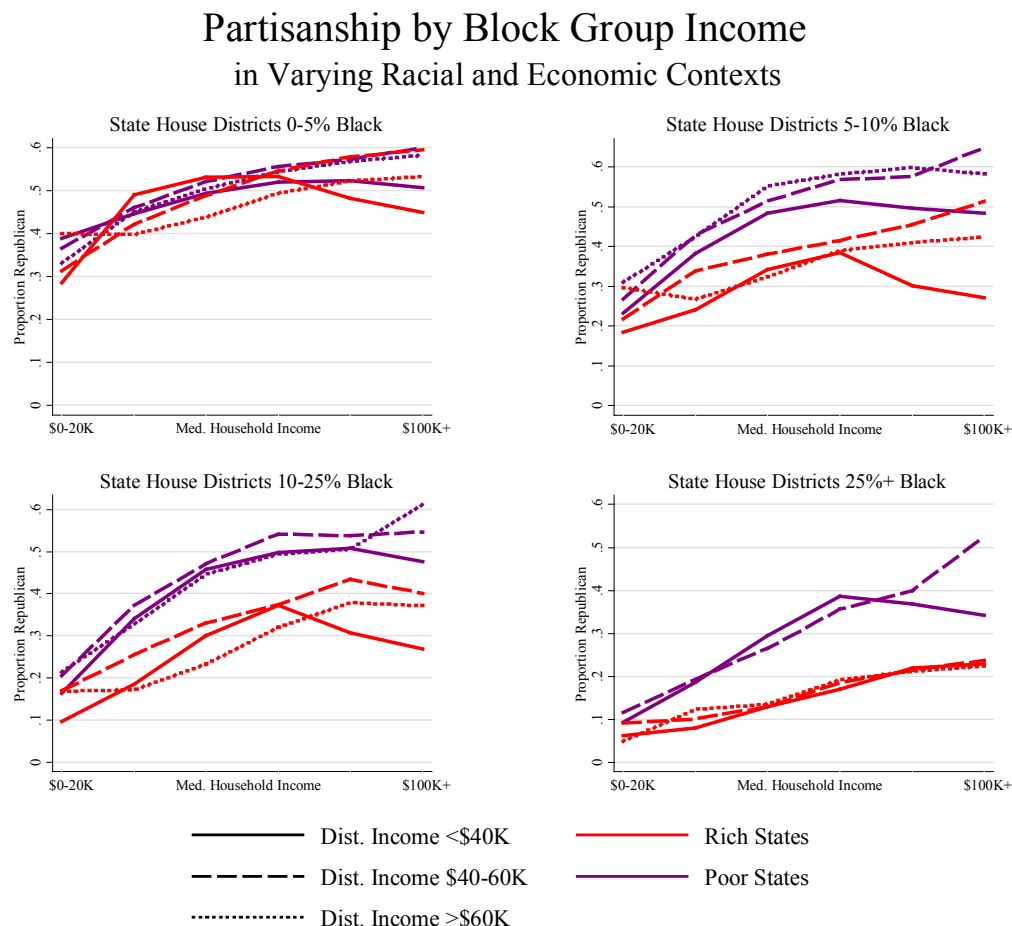
appear in the graphs, except in districts with large proportions of African-Americans.

These results show that differences between rich and poor states can be explained by geographic variation below the state level: by behavior among rich voters in more black districts in rich states, versus those who live among black voters in poor states. Moreover, this finding is not merely a reframing of the well-known fact that nearly all African-American voters (and hence African-American areas) are Democratic (Gelman et al., 2008, 73). In the Appendix, Figure A-6 restricts the analysis to block groups that are less than 10% Black, and Figure A-7 restricts the analysis to block groups that are more than 90% non-Hispanic white. While the income-party intercept is higher in these graphs (due to the removal of the poorest block groups dominated by Black, Democratic voters) the income-party relationship is still strong in the racially diverse districts in “poor states” and weaker in the less racially diverse areas of poor states.

The huge gap between rich block groups situated in heterogeneous districts in the poor states and rich block groups in heterogeneous districts in richer states explains to a large extent the South/non-South difference in voting behavior and the “red state, blue state” paradox discussed in Gelman et al. (2008). Voters with block group incomes over \$100,000 in heavily black districts in rich states affiliate with the Republicans, on average, at just over 20 percent, while voters in the same income group in poorer states affiliate with Republicans at twice that rate. While this paradox is an accurate statement of fact, states are not the “spatial regime” governing political behavior. Interstate differences in the income-party relationship rest almost entirely on differences in behavior in substate regions with substantial black populations and the absence of any impact of state-level income or district-level income on behavior in homogeneously white districts.

While individual-level nonparametric analyses summarize the income-party relationship with few modeling assumptions, they provide data for only 29 states, omit geographic labels, and contain empty cells. We address this limitation in two ways. First, in the Online Appendix, we replicate Figure 3 for each of the 29 party registration states, reporting the population sizes in each district-level/income-context income bin. Second, we address the empty-cells problem through multi-level modeling using both the voter registration data and precinct-level data (Ansolabehere and Rodden, 2012).

Figure 3: The Income-Party Relationship Depends on District-Level Racial Context, Not State Income



Note: Figures are derived from the full population of registered Democrats and Republicans in party registration states. N= 73,170,970.

### 3.4 Modeling Substate Geographic Heterogeneity in Income-Based Voting

We now employ a multi-level model that partially pools across districts to provide better estimates of the income-party relationship in regions where population data are inherently sparse within cells. We present these analyses in two ecological regressions. The first uses voter-population-weighted frequencies of two-party identifiers within block groups in the 29 voter-registration states to estimate the expected partisan gap between the richest and poorest income categories, by district. The second model uses 2008 precinct-level election returns to estimate geographic variation in the effect of income on the two-party



vote.

### 3.4.1 Multi-Level Model Using Catalist Data

For each state house district, the Catalist database permits construction of counts of voters in each income group, by partisan registration. We use these frequencies in a hierarchical linear model in which the Republican proportion of registrants in each block group income bin in each district is weighted by the total number of two-party registrants in each district income bin. We then regress two-party Republican registration on income, allowing the intercept and the coefficients on the six-category income variable to vary by district:

$$y_{ij} = \alpha + \beta_2 x_{2i} + \dots + \beta_6 x_{6i} + (\delta_{2j} x_{2ij} + \dots + \delta_{6j} x_{6ij}) + (\gamma_{2j} x_{2i[j]} + \dots + \gamma_{6j} x_{6i[j]}) + \epsilon_{ij}$$

where  $y_{ij}$  represents expected Republican share of the two-party vote in block-group  $i$  situated in district  $j$ ,  $\alpha$  represents an overall intercept term, and  $\beta_2 \dots \beta_6$  represent fixed effects at each ascending income category, adjusting for the estimated district-level random effects. The categorical income variable's omitted base category is block-group income of \$0 to \$20,000. The random intercept,  $\delta_j$ , is estimated by partially pooling the mean of observations  $i$  located in district  $j$  to the global mean intercept. We are most interested in the random-effects coefficients on the categorical income variables. These similarly vary by district  $j$  and are represented by the term  $\gamma_{kj}$  for each income category  $k = 2 \dots 6$ , where the base category is the bottom income category. The gamma coefficient on the top income category (for voters in block groups with income of \$100,000 or above) can be interpreted as the district-specific differential in the gap between the richest and poorest block groups. Higher positive (negative) values indicate stronger (weaker or more negative) gap in Republican identification between voters in the poorest and richest block groups of each district than would be expected otherwise. Finally, the error term,  $\epsilon_{ij}$  is assumed to be independently and identically distributed in each district  $j$ . The random effects and errors are each assumed to be normally distributed with constant mean and variance in each income category  $k$  (Gelman

and Hill, 2007, 258):

$$\delta_{kj} \sim \mathcal{N}(\mu_{\delta_k}, \sigma_{\delta_k}^2)$$

$$\gamma_{kj} \sim \mathcal{N}(\mu_{\gamma_k}, \sigma_{\gamma_k}^2)$$

$$\epsilon_{ij} \sim \mathcal{N}(\mu_{\gamma_j}, \sigma_{\gamma_j}^2)$$

This model is estimated using likelihood-based approximation to a Bayesian hierarchical model using the `R lmer` package (Bates, Maechler and Bolker, 2011), a choice widely adopted by scholars performing multilevel regression within states (e.g., Lax and Phillips 2009).

We apply a similar varying-intercept, varying-slope model to estimate the income-party relationship using the HEDA precinct data clustered within state house districts. The Republican share of the 2008 two-party presidential vote is regressed on precinct-level income, defined as the average of the block-group level median household income values in each precinct. This model allows the relationship between precinct-level income and the two-party vote to vary within state house districts. The hierarchical model for the income-party relationship for precinct  $i$  within district  $j$  is

$$y_{ij} = \alpha + \beta x_i + \delta_j + \gamma_j x_{i[j]} + \epsilon_{ij}$$

where  $\alpha$  is a general intercept term,  $\beta$  is a general coefficient on  $x_i$ , the precinct-level median household income variable expressed in tens of thousands of dollars (after accounting for district-level random effects), and  $\delta_j$  and  $\gamma_j$  are, respectively, the random intercept and income coefficients for precincts in each district  $j$ . We focus on the term  $\gamma_j$ , which is analogous to an interaction term between income variable and the district dummy variable.

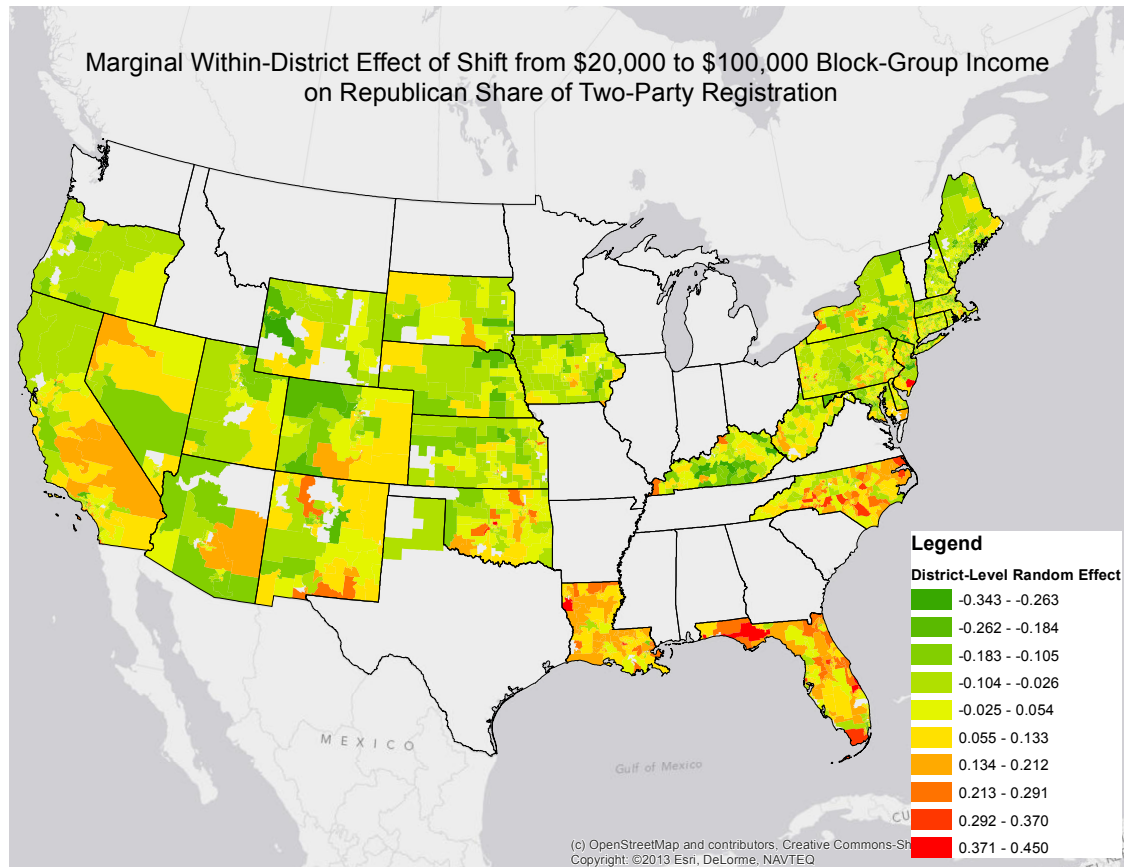
Rather than displaying these methods using non-spatial methods such as confidence-interval plots, we present the estimates using a national map of state house districts. For each model, the  $\gamma_j$  random income effects estimated for each grouping geography  $j$  are presented as a heat map graduated by twelve equal intervals. Districts in which the marginal district-level income-partisanship relationship is lower than expected appear at the green end of the scale, those where it is weaker appear at the redder end of the scale.

Figure 4 demonstrates the importance of sub-state geography to the income-party relationship. The map presents  $\gamma_6$ , the random effect coefficient on the indicator for registered voters in block groups with incomes above \$100,000 per year, relative to a base category of households earning \$20,000 per year or less. It thus captures each district's rich-poor gap in Republican identification among two-party registrants, relative to a national baseline. These random effects range from -0.34 to 0.45, indicating that the gap is as many as 34 points lower or 45 points higher than expected.

Several anomalies stand out from this map. Large swaths of the party-registration states of the Old South—North Carolina, Florida, and Louisiana—have a stronger income-party gap than would otherwise be expected, especially (but not only) in the “Black Belt” areas of those states. For example, the relationship is especially strong in the old cotton cultivation areas of eastern North Carolina, and much weaker in the western part of the state. Similarly, the ethnically diverse and economically polarized areas of California's Central Valley region, an area dominated by agribusiness, have a strong relationship between income and party. As expected, the slope is lower than would otherwise be predicted in most metropolitan areas, especially in districts of the Northeast megalopolis running from DC to Boston that have been the subject of “blue-state voter” stereotyping. However, such differences are far less pronounced in the non-metropolitan areas of these states. Something that would not be discovered only in “red-state, blue-state” stereotyping is the strong cluster of lower-than-expected income-based voting in Utah, Idaho, Northern Arizona and Nevada—the so-called “Mormon Corridor.” In these areas, voters in both rich and poor precincts are strong Republican party identifiers, creating a large cluster across the Mountain West in which the income-party relationship is substantially flatter than elsewhere. Even within these states, however, there are deviations. For example, the income-party relationship is much stronger in more racially and religiously diverse areas of Salt Lake City. These patterns would not be detected either with widely used county-level data (Morrill, Knopp and Brown, 2007, 2011) or geocoded survey data.

Vote-based results are largely in agreement with party-registration results, with previously well-known substate political regions measurable in sharper relief and with greater detail. Figure 5 displays a map of the random slope coefficient,  $\gamma_j$ , which captures the district-level deviation of the income coeffi-

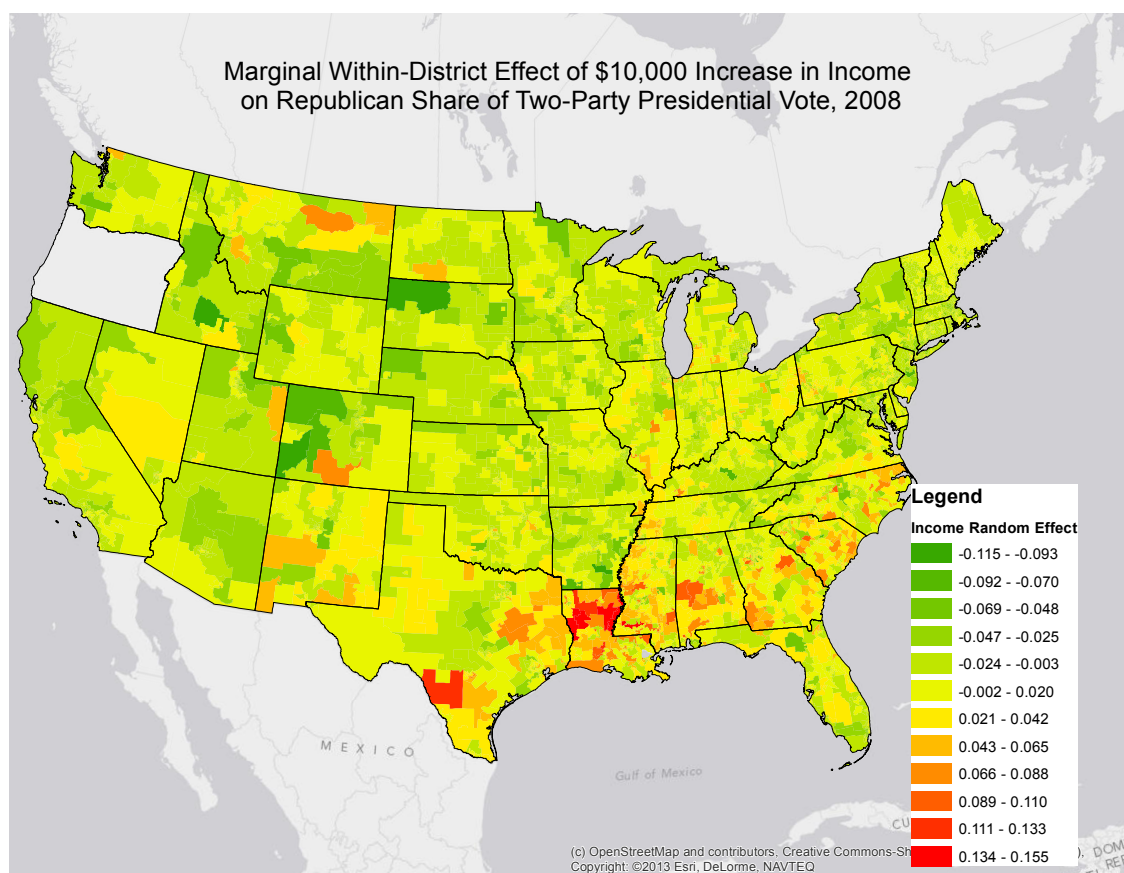
Figure 4: Income-Based Voting is Stronger Than Expected in Poor Rural Districts



Map of estimates of the marginal effect of income on individual Republican party registration, using data on Republican and Democratic registration in block-group income categories in party-registration states. Increasing shades of green indicate a less positive marginal district-specific effect between income and party, while increasing shades of red indicate a more positive income-party relationship than would be predicted otherwise.

cient from the expected value. The random-effects coefficient on income (in tens of thousands of dollars) ranges from -0.174 to 0.171, indicating a place that is 17 points less Republican than expected for every \$10,000 in per capita income, to a place that is 17 points more Democratic than expected. The Black Belt along the Mississippi River and Black Belt appears prominently in this graph, as does California's Central Valley and predominantly Latino Rio Grande Valley.

Figure 5: Income-Based Voting is Stronger in State House Districts with Rural Minority Poverty



Map of estimates of the district-level random effect of income on the 2008 McCain presidential vote, based on data from the Harvard Election Data Archive. Increasing shades of green indicate a lower income effect than expected, while increasing shades of red indicate a higher effect than expected.

### 3.4.2 LISA Analysis Detects Geographic Clusters of Deviant Income-Based Voting Linked to Race and Religion

We formalize the detection of sub-state clustering of anomalous income-based voting by calculating local indicators of spatial autocorrelation (LISA) statistics using the random effect coefficients from the precinct-level voting analysis. Such methods have been used before to identify substate regions, but typically using coarser county-level data, and not modeling within-county or within-district preferences. We plot local spatial autocorrelation of the income random effects estimated in Section 3.4.1.

The Local Moran’s  $I$  is calculated for each district  $i$  as the correlation of the income random effect in unit  $i$  and its adjacent neighbors (indexed by  $j$ , with  $J$  total neighbors) (Anselin, 1995):

$$I_i = (X_i - \bar{X})w_{ij} \sum_j (X_j - \bar{X}) \quad (1)$$

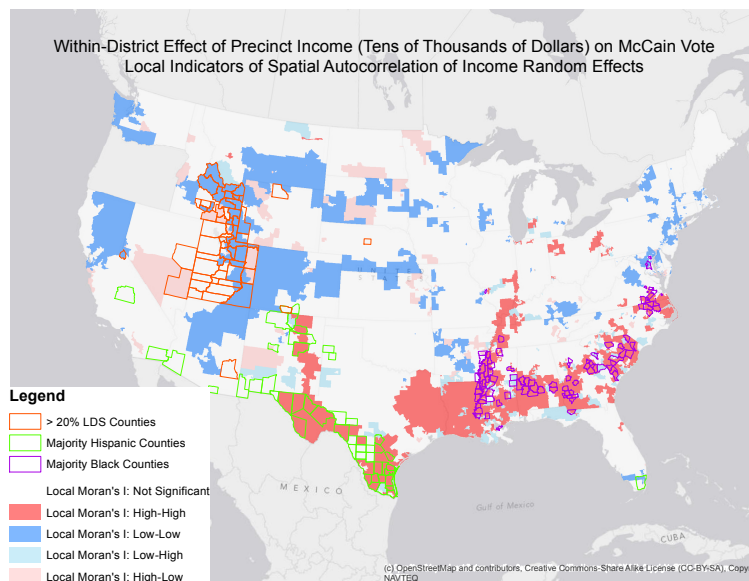
where  $w_{ij}$  represents weights coded 1 if the units  $i$  and  $j$  are contiguous and 0 otherwise.<sup>5</sup> Local positive and negative autocorrelation is represented by a standard coding scheme assigning four categories: high-high (districts with higher-than-average income random effects adjacent to other districts with higher than average income random effects), low-low (lower-than-average random effects adjacent to lower-than-average random effects), low-high (lower-than-average surrounded by higher-than-average) and high-low (higher-than-average surrounded by lower-than-average). Following standard data visualization techniques, these values are color-coded if they are statistically significant at the  $p < 0.05$  level, calculated using permutation-based inference in which outcome variables are randomly assigned across units to define a null distribution (Anselin, 2003).

Figure 6 displays a map of  $z$ -statistics associated with each district’s Local Moran’s  $I$ . The graph presents in sharper relief the natural geographic regions in which the income-party relationship is stronger (or weaker) than expected. As expected, the majority Black areas of the Deep South and surrounding counties are classified into clusters of “high-high” values. Other areas of poor Southern states, such as Northern Alabama and East Tennessee that lack a history of racialized poverty, do not fall in these

---

<sup>5</sup>The weights matrix was generated using “queen contiguity” in which all adjacent units, even if connected only at one vertex, is coded “1” in the spatial weights matrix.

Figure 6: The Income-Party Relationship is Uniformly Strong in the Deep South Black Belt and Rio Grande Valley, Weak or Inverted in the Northeast, Mormon Corridor and Great Plains Bible Belt



Local Moran's  $I$  for the income slope random effect, with overlay of majority-black, majority-Hispanic, and Mormon-dominated ( $> 20\%$  LDS) counties.

clusters. Another band of counties roughly running along the Rio Grande River is identifiable as a place with large Hispanic populations. Similarly, in other clusters of the country precinct-level income-party relationship is lower than expected. In the Northeast, differences in voter behavior associated with states are driven by unusual income-based voting behavior in the metro areas of the Northeast Corridor. In the Mountain West, the “Mormon Corridor” consisting of homogeneously Republican religious adherents at all income levels, runs from the Eastern Idaho into Northern Arizona and Western Colorado, and is matched by less prominent clusters of Evangelical Christians across the Great Plains, including in (notorious) Kansas. The substate sectionalism (Gimpel and Schuknecht, 2002*b*) of California is also obvious here: in northern California (the “bluest” part of a blue state) there is a uniformly flat relationship between income and party, while the relationship in southern California looks much like the national average.

## 4 Discussion

What is the right way to study the relationship between income and voting? Whether we should pay attention to covariation within states, versus covariation within districts, depends to a considerable extent on the nature of research questions. Examining the link between income and partisanship at the state level is only relevant if we care about relationships within states. Even then, the data generating process underlying these relationships has its roots in substate political geography. While voters in wide swaths of the country behave quite similarly (especially those in homogeneously white precincts), voters in clearly identifiable behavioral clusters act differently. The relationship between block-group income and partisanship is stronger than expected in the Black Belt and Rio Grande, while elsewhere in states containing these regions the patterns of income-based voting look like the rest of the country. Meanwhile, red-state/blue-state stereotyping belies underlying variation within blue states, with prominent clusters of relatively uniform Democratic registration and voting across all income groups in Northern California, the Northeast Corridor, and in smaller pockets of other metropolitan areas, but less uniformly outside metro areas. It also belies substantial differences across “red states,” with the conventional wisdom on weaker income-based voting among the devout holding in the Mormon Corridor and parts of the Great



Plains Bible Belt.

While our findings are, for the most part, consistent with the individual-level estimation in Gelman et al. (2008), we show that the estimation of state-level random effects misses important “sub-state sectionalism” in voter behavior (Gimpel and Schuknecht, 2002*a*, 2004) that is not detectable in non-spatial individual-level data or even by subsetting data demographically within states. Beginning with statistics aggregated at the state-house-district level, then analyzing block-group and precinct-level aggregate behavior within districts within states, we show that behavior clusters within sets of areas that do not neatly follow state lines yet follow clear historical boundaries.

Our analysis shows that by compiling large-scale records from the Census Bureau, voter registration system, and election archives, we can detect meaningful clustering in voter behavior without adapting our analyses to data that are the “wrong size”: not representative of underlying geographic fractionalization, or determined by convenience rather than by voter behavior. We demonstrate that inferences about voter behavior based only on state-level voter behavior are suitable for determining state-level opinion but do not align with behavioral communities of interest. Limiting geographic comparisons of income-based voting to cross-state comparisons, even when accompanied by “deep interactions” (Ghitza and Gelman, 2012), assumes that data are generated from a single data generating process in each state. Below state lines, methods such as MRP typically improve the accuracy of estimates only marginally (Tausanovitch and Warshaw, 2013). However, political behavior covaries within clusters that are readily identifiable in low-level aggregate data sets.

## **4.1 Concerns and Next Steps**

Readers might note that our critique takes on the modifiable areal unit problem, but that our analyses are also vulnerable to this problem. Since all studies that rely on any choice of geography are susceptible the MAUP, we take the following precautions. First, we verified through the analysis of local spatial autocorrelation that the geographic clustering of income-based voting occurs without respect to state boundaries, and we show that clustering of behavior commonly occurs across multiple adjacent districts, reducing the likely influence of redistricting (i.e., zoning) on the analysis. Second, the geographic clustering of

income effects in the Black Belt and other cluster areas are often observed in aggregate county data as well, indicating that they are not an artifact of redistricting (see Appendix). Third, even if a portion of the effects here result arise from the redistricting process, districts themselves are areas of interest. Policy is made by legislators elected within these particular sets of boundaries.

Another potential concern may relate to differences between voter-registration and two-party voting, as well as the differences between the district-level results based on registration and those based on party support. One reason for the difference is that voting-based results adopt a linear functional form assumption with aggregate precinct data, rather than weighted block-group-level counts. Some of the cells in the voter-registration data are extremely sparse, and this is especially true of the cells in the top income categories in poor, non-white districts. The party registration random effects are calculated with respect to the 29 party-registration states only, so the random effects estimated for the two groups will be slightly different.

While party registration tends to be a lagging indicator of partisanship, we show that it correlates so strongly with other measures of partisanship that it is a suitable proxy for self-identification-based measures. In the Appendix, we examine the relationship between registration and other partisanship measures, including survey-based party identification and vote choice.<sup>6</sup> We also examine the relationship between block-group level and individual level income and district level partisanship and district representation in state houses. We also analyses using counties, rather than state house districts, as the geographic cluster of interest.

A natural suggestion arising from our results is that the data we analyze could be incorporated into existing methods developed to extend small-area estimation using survey data. While it would be outside the scope of this paper, incorporating information from voter lists and block-group level demographics would likely improve the accuracy of “deep interactions” using heavily interacted and post-stratified models (Ghitza and Gelman, 2012). Currently, such methods assume homogeneity within subgroups, and occasionally adjust by aggregate-level covariates in small geographies to improve prediction and

---

<sup>6</sup>While these analyses appear in Ansolabehere and Hersh (2012), those presented here are original to the article.

account for geographic heterogeneity. Data such as the Catalyst block-group level party frequencies would likely improve individual-level estimation within these small areas. Particularly in places with little data, where substantial shrinking to the global mean is necessary, block-group level estimates could be used as priors.<sup>7</sup> Jackson, Best and Richardson (2008), for example, show that ecological data adjusted using only a small number of individual-level observations (often, only 10 or 15 individuals) containing the same covariates in individual form, typically produces estimates with much lower mean square error than estimates based on disaggregated survey data.

## 4.2 Conclusion

Our analysis demonstrates the importance of choosing geographic units and data that are the suited to study local variation in political behavior. The application of simple methods and cognizance of the modifiable areal unit problem show that questions such as “What’s the Matter with Kansas” and “What’s the Matter with Connecticut” are imprecisely posed, whether the question is approached from the perspective of state-level data or in national data that make population-wide or regional inferences about individual-level voter behavior (Bartels, 2006). Many of our findings are consistent with past accounts in which American regions have been grouped without respect to state boundaries (Gimpel and Chinni, 2011; Garreau, 1991). Yet we show that these American subregions are also distinctive in the extent to which neighborhoods and state house districts either vote without respect to income class (as in the Mormon Corridor) or in line with local race and class divisions (as in the Black Belt).

Both “red states” and “blue states” have heterogeneous internal “spatial regimes” (Anselin, 1995). The evidence for these regimes appears, to an extent, in county-level election maps, notably where voters of all income levels lean toward one party or the other. But no previous analysis has provided precise estimates of income and Republican partisanship based on income and voting data well below the state level.

---

<sup>7</sup>For a discussion of models integrating individual-level and aggregate data, see Greiner and Quinn (2010), Glynn et al. (2009), Jackson, Best and Richardson (2006), and Jackson, Best and Richardson (2008).

In our analysis, we presented just one example entailing block-group-level and precinct-level covariance between income and partisanship. This finding, and others using micro-scale ecological data, hint at the plausible resurrection of ecological research in line with Key (1949), but with finer-grained data. Consider all the variables that are available from the Census Bureau at the Census block group level: income and education levels, family composition, immigrant populations, military involvement, work and commuting characteristics, housing and public assistance statistics. These variables can be connected with political measures of party registration, voting history, and other individual data coming from the registration system as well as with precinct data. This linking of records permits characterization of local communities without modeling assumptions and without stretching nationally representative (but locally unrepresentative) survey data to fit geographic analysis. Moreover, because electoral data are a matter of public record, they permit reporting of behavior at levels well below what is typically permitted in anonymous (and respondent-protecting) survey data.

Given our findings, future research that examines the micro-foundations of income-based voting may need to engage more consciously with theories that explain racial threat, race-based sorting, and racial liberalism within sub-state regions (e.g. Tesler and Sears 2010, Acharya, Blackwell and Sen 2013) . They may account for geographic differences using increasingly available political “big data.” With such data we are able to ascertain the regions in which voters of similar types behave in concert. These regions frequently happen to be defined not by state lines, but by communities of shared racial, ethnic, and religious interest.

## References

- Abrams, Samuel and Morris Fiorina. 2012. “‘The Big Sort That Wasn’t’: A Skeptical Reexamination.” *PS: Political Science and Politics* 45(2):203–210.
- Acharya, Avidit, Matt Blackwell and Maya Sen. 2013. “The Political Legacy of American Slavery.”  
**URL:** <http://www.mattblackwell.org/files/papers/slavery.pdf>
- Alesina, Alberto and Edward Glaeser. 2004. *Fighting Poverty in the United States and Europe: A World of Difference*. Oxford: Oxford University Press.
- Anselin, Luc. 1995. “Local Indicators of Spatial Association-LISA.” *Geographical Analysis* 27(2):93–115.
- Anselin, Luc. 2003. *GeoDa 0.9 User’s Guide*. Urbana: University of Illinois. Downloaded 16 June 2013.
- Ansolabehere, Stephen and Eitan Hersh. 2012. “Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate.” *Political Analysis* 20(4):437–459.
- Ansolabehere, Stephen and Jonathan Rodden. 2011. “Harvard Election Data Archive.” online.  
**URL:** <http://projects.iq.harvard.edu/eda>
- Ansolabehere, Stephen and Jonathan Rodden. 2012. “Harvard Election Data Archive.” online.  
**URL:** <http://projects.iq.harvard.edu/eda>
- Bartels, Larry M. 2006. “What’s the Matter with What’s the Matter with Kansas?” *Quarterly Journal of Political Science* 1:201–226.
- Bates, Douglas, Martin Maechler and Ben Bolker. 2011. “lme4 Package.” Online.
- Blumer, Hubert. 1948. “Public Opinion and Public Opinion Polling.” *American Sociological Review* 13(5):542–549.

- Butler, Daniel M. and Eleanor Neff Powell. 2012. "Understanding the Party Brand." Working paper.
- Cho, Wendy Tam, James G. Gimpel and Iris S. Hui. N.d. "Voter Migration and the Geographic Sorting of the American Electorate." *Annals of the Association of American Geographers*. Forthcoming.
- Cho, Wendy Tam and Neil Baer. N.d. "Environmental Determinants of Racial Attitudes Redux: The Critical Decisions Related to Operationalizing Context." . Forthcoming.
- Dawson, Michael C. 1995. *Behind the Mule*. Princeton: Princeton.
- Ellen, Ingrid Gould. 2000. *Sharing America's Neighborhoods: The Prospects for Stable Racial Integration*. Harvard University Press.
- ESRI. 2008. U.S. Census Block Groups. In *StreetMap North America*. ESRI.
- Farley, R, C Steeh, M Krysan, T Jackson and K Reeves. 1994. "Stereotypes and Segregation: Neighborhoods in the Detroit Area." *American Journal of Sociology* pp. 750–780.
- Fitch, Catherine A and Steven Ruggles. 2003. "Building the National Historical Geographic Information System." *Historical Methods* 36(1):41–51.
- Fotheringham, A.S. and D.W.S. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." *Environment and Planning A* 23(7):1025–1044.
- Frank, Thomas. 2004. *What's the Matter With Kansas? How Conservatives Won the Heart of America*. New York: Henry Holt and Co.
- Garreau, Joel. 1991. *Edge City: Life On the New Frontier*. New York: Doubleday Books.
- Gelman, Andrew, Boris Shor, Joseph Bafumi and David Park. 2007. "Rich State, Poor State, Red State, Blue State: What's the Matter With Connecticut?" *Quarterly Journal of Political Science* 2(345–367).
- Gelman, Andrew, Boris Shor, Joseph Bafumi, David Park and Jerome Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton, NJ: Princeton.

- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Gelman, Andrew and Thomas Little. 1997. "Poststratification Into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2):127–135.
- Ghitza, Yair and Andrew Gelman. 2012. "Deep Interactions With MRP".
- Gibson, Keegan and Daniel Gleason. 2012. "Obama's WV Problem Stretches Into PA." *Politics PA* . Accessed August 12, 2012.  
**URL:** <http://www.politicspa.com/obamas-wv-problem-stretches-into-pa/36833/>
- Gimpel, James and Dante Chinni. 2011. *Our Patchwork Nation: The Surprising Truth About the "Real America"*. New York: Gotham.
- Gimpel, James G. and Jason E. Schuknecht. 2002a. "Political and Demographic Foundations for Sectionalism in State Politics: The Connecticut Case." *American Politics Research* 30:193–214.
- Gimpel, James G. and Jason E. Schuknecht. 2002b. "Reconsidering Political Regionalism in the American States." *State Politics and Policy Quarterly* 2(4):325–352.
- Gimpel, James G and Jason E Schuknecht. 2004. *Patchwork Nation: Sectionalism and Political Change in American Politics*. Ann Arbor: Michigan.
- Glaeser, Edward L. and Bryce A. Ward. 2006. "Myths and Realities of American Political Geography." *The Journal of Economic Perspectives* 20(2):119–144.
- Glynn, Adam, Jon Wakefield, Mark S. Handcock and Thomas S. Richardson. 2009. "Alleviating Ecological Bias in Voter Turnout Models (and other Generalized Linear Models) with Optimal Subsample Design.".
- Grannis, Rick. 2009. *From the Ground Up: Translating Geography into Community through Neighbor Networks*. Princeton, NJ: Princeton University Press.

- Greiner, D. James and Kevin Quinn. 2010. "EXIT POLLING AND RACIAL BLOC VOTING: COMBINING INDIVIDUAL-LEVEL AND RC ECOLOGICAL DATA." *Annals of Applied Statistics* 4(4):1774–1796.
- Jackson, Christopher, Nicky Best and Sylvia Richardson. 2006. "Improving Ecological Inference Using Individual-Level Data." *Statistics in Medicine* 25:2136–2159.
- Jackson, Christopher, Nicky Best and Sylvia Richardson. 2008. "Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors." *Journal of the Royal Statistical Society: Series A* 171(1):159–178.
- Key, V O. 1949. *Southern Politics in State and Nation*. New York: Vintage.
- King, Gary. 1997. *A Solution to the Ecological Inference Problem*. Princeton University Press.
- Lax, Jeffrey and Justin Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1):107–121.
- McCarty, Nolan, Keith Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.
- McGhee, Eric and Daniel Krimm. 2009. "Party Registration and the Geography of Party Polarization." *Polity* 41(3):345–367.
- Morrill, Richard, Larry Knopp and Michael Brown. 2007. "Anomalies in Red and Blue: Exceptionalism in American Political Geography." *Political Geography* 26:525–553.
- Morrill, Richard, Larry Knopp and Michael Brown. 2011. "Anomalies in Red and Blue II: Towards an Understanding of the Roles of Setting, Values, and Demography in the 2004 and 2008 U.S. Presidential Elections." *Political Geography* 30.
- Openshaw, Stan and P.J. Taylor. 1981. The Modifiable Areal Unit Problem. In *Quantitative Geography: A British View*, ed. N. Wrigley. Routledge & Kegan Paul Books pp. 60–69.



- Park, David, Andrew Gelman and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Post-stratification: State-Level Estimates from National Polls." *Political Analysis* 12:375–385.
- Peterson, Paul E. 1981. *City Limits*. Chicago: Chicago.
- Rao, J.N.K. 2003. *Small-Area Estimation*. Hoboken, NJ: Wiley and Sons.
- Reardon, Sean and Kendra Bischoff. 2011. "Income Inequality and Income Segregation." *American Journal of Sociology* 116(4):1092–1153.
- Scott, James C. 1998. *Seeing Like a State*. New Haven: Yale University Press.
- Scott, James C. 2010. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven: Yale University Press.
- Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2):330–342.
- Tausanovitch, Chris and Jonathan Warshaw. 2011. "Testing Theories of Representation in Congress, State Legislatures and Cities Using a 'Super Survey'."
- Tesler, Michael and David O. Sears. 2010. *Obama's Race: The 2008 Election and the Dream of a Post-Racial America*. Chicago: University of Chicago Press.
- United States. Bureau of the Census. 2006. State Legislative Districts. Technical report Washington, DC: . Downloaded 15 December 2012.
- URL:** [http://www.census.gov/geo/www/cob/sld\\_metadata.html](http://www.census.gov/geo/www/cob/sld_metadata.html)
- Vigdor, Jacob. 2006. "Fifty Million Voters Can't Be Wrong." *NBER Working Papers* (12371).
- Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Measure District-Level Opinion on Individual Issues?" *Journal of Politics* 74(1):203–219.

Wong, Cara, Jake Bowers, Tarah Williams and Katherine Drake Simmons. 2012. "Bringing the Person Back In: Boundaries, Misperceptions, and the Measurement of Racial Context." *Journal of Politics* 74(4):1153–1170.

# Supporting Information

<b>A</b>	<b>Journals Used in Multi-Field Literature Review on the MAUP</b>	<b>34</b>
<b>B</b>	<b>Party Registration as a Measure of Partisanship</b>	<b>34</b>
<b>C</b>	<b>Validating Block-Group Income Measure Using CCES Data</b>	<b>41</b>
<b>D</b>	<b>OLS Regression Analysis of Income-Party Relationship</b>	<b>41</b>
<b>E</b>	<b>Disaggregating Nonparametric Results by State</b>	<b>43</b>
<b>F</b>	<b>Replication of Figure 3 in for Non-Black Block Groups</b>	<b>46</b>
<b>G</b>	<b>Replication of Figure 3 with Precinct Returns</b>	<b>48</b>
<b>H</b>	<b>Summary Statistics for Multi-Level Models</b>	<b>49</b>
<b>I</b>	<b>Details on Multi-Level Modeling Using the lme4 Package</b>	<b>50</b>
<b>J</b>	<b>Variation in Precinct-Level Income-Based Voting Within Counties</b>	<b>53</b>
<b>K</b>	<b>State-by-State Analysis of Party Registration, Block-Group Income, District Income, and District Percent Black</b>	<b>54</b>

This Appendix contains a series of figures, tables, and methodological discussions referenced in the text.

## **A Journals Used in Multi-Field Literature Review on the MAUP**

Using Google Scholar, we searched the literatures in three disciplines—political science, geography, and statistics—for a discussion of three related search terms: “modifiable areal unit problem,” “ecological inference problem,” and “contextual effects.” This search includes all articles that appear on Google Scholar, but may omit other articles that were not found by the search engine. An article need only mention the term once to be included in the article count.

Journals from the following fields were searched:

- Geography: *Annals of the Association of American Geographers*, *Journal of Economic Geography*, *Progress in Human Geography*, *Transactions of the Institute of British Geographers*
- Political Science: *American Journal of Political Science*, *American Political Science Review*, *American Politics Research*, *Journal of Politics*, *Political Analysis*, *Quarterly Journal of Political Science*.
- Statistics: *Annals of Statistics*, *Biostatistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society, Series A*, *Journal of the Royal Statistical Society, Series B*, *Statistical Science*.

## **B Party Registration as a Measure of Partisanship**

The dependent variable throughout our study has been an estimate of voter partisanship. In the main analysis, we measure partisanship based on party registration data. Since party registration is neither the only basis nor the most common basis for studying partisanship, the validity of the measure merits some attention. There are essentially three common ways to study the partisanship of voters: one based on vote choice, a second based on self-reported identification with political parties (i.e. party ID), and a third based on registration status. Each of these three measures - vote choice, party ID, and party registration

- has advantages and disadvantages for researchers. In this section, we demonstrate that while these measures differ slightly, in recent years they have been highly correlated and each measure an aspect of underlying partisan identity.

Vote choice represents the clearest translation of voter preferences into political outcomes. However, political scientists have studied party ID instead of vote choice in large part because party ID captures more stable partisan preferences than voters' decisions in any one election. As McCarty, Poole and Rosenthal (2006) argue in support of using party ID, the measure is "less influenced by election-specific factors" (75). But party ID has drawbacks too. Since it can only be estimated through surveys, the measure may be corrupted by sample selection and misreporting biases, and national surveys will not be useful for studies of small sub-populations. Party registration is advantageous in that it is available at the individual-level in public records; however, it is only available in three-fifths of the states that collect the data. Fortunately, the states that have party registration are quite representative of the entire nation, as Abrams and Fiorina (2012) and McGhee and Krimm (2009) discuss.

The theoretical motivation for studying vote choice and party ID have been well established in the literature. The basis for studying party registration is not as well established. Thus we offer three justifications for our use of party registration data as a measure of partisanship. Our first justification is that it is independently an interesting measure worthy of our attention. Our second justification is that party registration is very closely related to vote choice and party ID both at the individual-level and aggregate levels. This is especially apparent in the case of party identification which is nearly perfectly collinear with party registration. Our third justification is seen in a robustness check by which we replicate with individual-level survey responses one of the graphs we made using party registration and census income data.

Party registration is, in itself, an important phenomenon in American politics. In many states, party registration is a pre-requisite for voting in primaries. Therefore the population of registered party affiliates represents the universe of the core voters who determine the party nominees. Furthermore, in every state with party registration, these data provide the basis for political mobilization. Campaigns utilize the public records to decide which voters and which geographic areas merit their attention. Thus, the

connection between party registration and other measures of partisanship aside, it is useful to understand the variance in party registration across voters and across places.

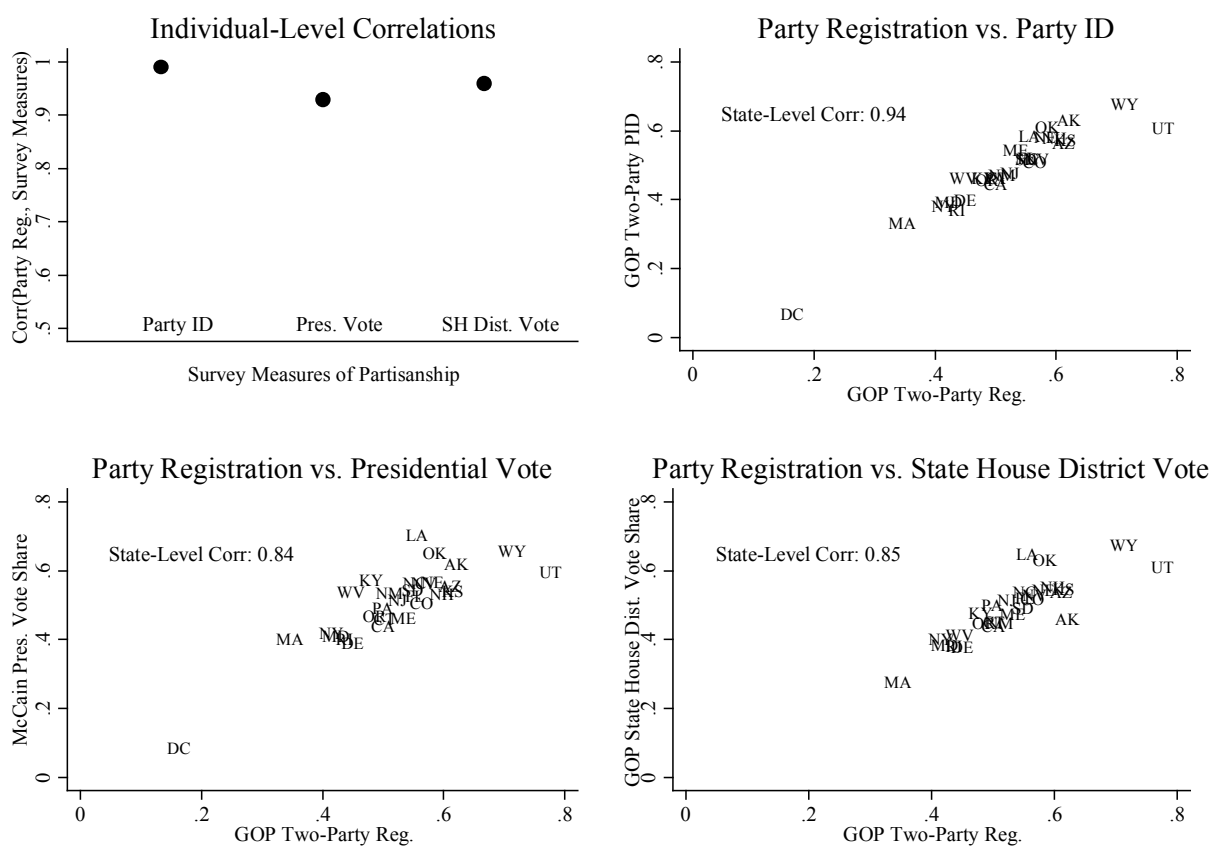
Our second justification is that in addition to the independent virtues of party registration, the measure is also very highly correlated with vote choice and party ID. To show this, we first rely on the 2008 Cooperative Congressional Election Study (CCES) that has previously been merged with public records of party affiliation (Ansolabehere and Hersh, 2012). Among respondents who are registered Republican according to public records, 96% of them identify as Republican, 89% reported voting for John McCain for President, and 91% said they voted for the Republican candidate for state representative. For respondents who are registered Democratic, 95% identify as Democratic, 87% reported voting for Barack Obama, and 95% reported voting for the Democratic candidate for state representative. Notice there is slightly more defection for Presidential voting compared to down-ballot voting, which is not surprising given that voters have much less information about non-Presidential races and therefore rely more heavily on partisanship for their vote choice in these contests.

Figure A-1 shows the state-by-state relationships between party registration stemming from public records and survey-based measures of partisanship. The first sub-plot simply shows the individual-level tetrachoric correlations between party registration and each of the survey measures. The second sub-plot shows the near perfect collinearity between party registration and party ID. Only one state, Utah, is slightly off-diagonal in this graph. To the extent that scholars value party ID as a measure of partisanship, the individual-level and state-level relationships on display indicate that party registration is a valid proxy for party ID.

In addition to studying party ID, we investigate the relationship between party registration and self-reported vote choice in the 2008 Presidential race and in the respondent's state house district election. At the state level, the relationship between party registration and vote choice is very tight - especially with respect to voting in the down-ballot races such as for state house district representative. We display statistics about voting for state representatives because most of our analysis in this manuscript investigates how partisanship varies according to state house district.

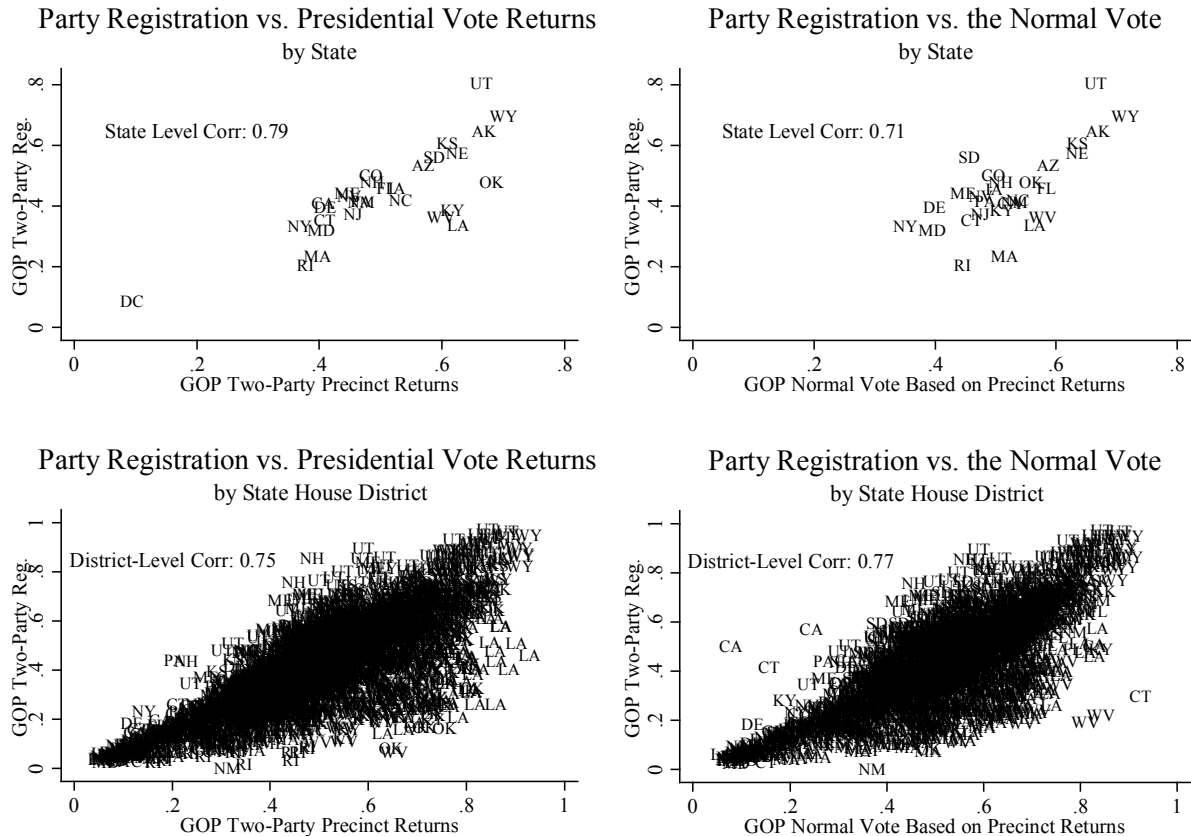
Aside from the survey analysis, we also compare party registration to actual vote returns. We do

Figure A-1: Party Registration Compared to Survey-Based Measures of Partisanship



Note: There are 9,621 weighted respondents with non-missing values on two-party PID and two-party registration. There are 9,743 respondents with non-missing values on two-party registration and two-party Presidential vote choice. And there are 8,421 weighted respondents with non-missing values on two-party registration and two-party state representative vote choice. Data source: 2008 CCES matched with Catalist voter registration data.

Figure A-2: Party Registration Compared to Precinct Returns



Note: Precinct returns in Oregon were not available for use in this graph. Additionally, DC precinct returns are not available for the subplots estimating normal vote. In the Catalist 1% sample, 691,345 registered Democrats and Republicans were matched by Census block group to the vote returns from their precincts. Data source: Catalist 1% analytics sample and Ansolabehere/Rodden precinct returns.



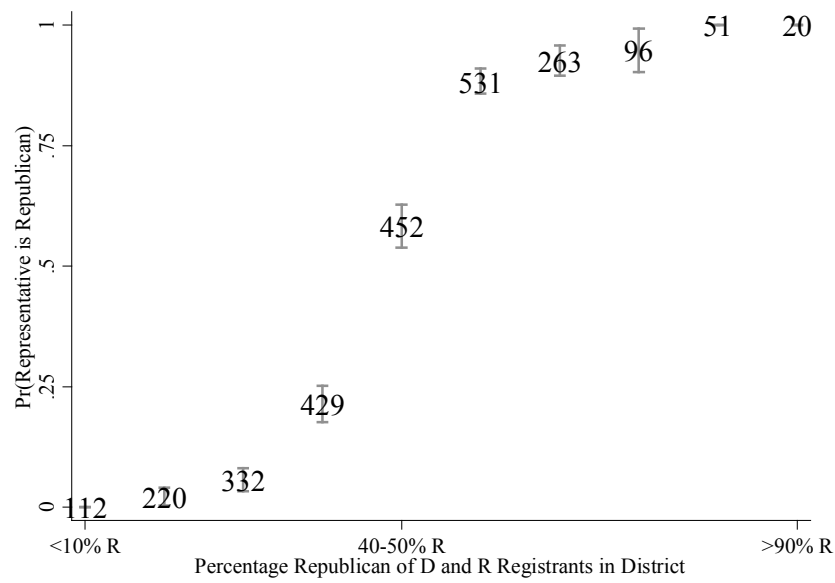
this by combining two sources of data. First, for a 1% sample of voters in Catalist’s database, we have a record of each registrant’s party affiliation and the census block group of his or her residence. Second, we utilize precinct-level vote returns collected by Stephen Ansolabehere and Jonathan Rodden. By merging these files based on block group, for 1% of all registered voters we have both their registration status and the vote distribution in the precinct in which they live.

In Figure A-2, we aggregate the registration and vote choice data to the state level and to the state house district level. We show the relationship by aggregating precinct-level Presidential vote returns from the 2008 election and also by aggregating a precinct-level “normal vote” calculated in Ansolabehere and Rodden (2011). The normal vote measure averages returns from down-ballot contests and from prior years, but the inputs to the measure vary based on data availability across states.

As with down-ballot voting measured on the survey, the normal vote measure ought to be a more stable signal of partisanship than voting returns from a single Presidential election. Indeed, the normal vote measure is more highly correlated with party registration than Presidential returns are in the state house district graphs. In the statewide graph, the correlation measure is lower for the normal vote only because we were not able to use data from Washington DC in the normal vote subplots. While both measures stemming from vote returns are clearly very similar to the party registration measure, here we see that there are several outlier states. Kentucky, West Virginia, Louisiana and Oklahoma vote somewhat more Republican than their registration records suggest. Nevertheless, notice that these states are not outliers in the survey comparisons. To the extent that party ID, party registration, and vote choice pick up on slightly different aspects of partisanship, the registration and ID measures are closer to one another and are both slightly different from vote choice in some states.

As another look at the robustness of party registration data, we rely on data shared by Daniel Butler that indicates the party affiliation of each state legislator elected in the 2010 midterm election (see Butler and Powell (2012)). We merge Butler’s data with the proportion of registered Republicans in each district, as derived from the Catalist estimates. In Figure A-3, the Republican vote share is binned into 10-percentage point groupings. For instance, the first grouping includes all districts that are 0-10% Republican. Each point is represented by the number of districts falling into each bin. The figure

Figure A-3: Party Registration Compared to 2010 District Representatives



Note: Data on legislator partisanship provided courtesy of Daniel Butler. Districts in Nebraska are not included because their representatives are non-partisan. The numbers in the figure represent the number of districts that fall into each party-registration bin.

demonstrates the strong relationship between party registration and electoral outcomes in state house districts.

The close relationships demonstrated here among different measures of partisanship not only help to justify our reliance on the most under-utilized of these measures, but it also helps advance scholarship on partisanship in general. Partisanship as measured by party ID, registration, or vote choice, and measured in surveys or in public records, exhibits a remarkable degree of conformity. Each of these measures has virtues and disadvantages, and that is why we replicate our analysis using alternative measures. Nevertheless, future projects that will be unable to replicate their analyses with multiple measures of partisanship may gain confidence in their results from the comparisons illustrated here.

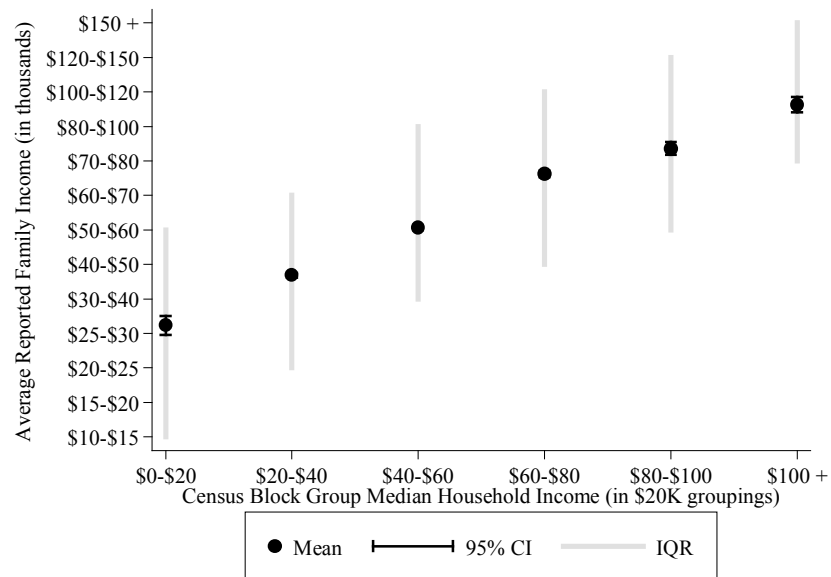
## **C Validating Block-Group Income Measure Using CCES Data**

We compare our income-based survey results to data from the 2008 Cooperative Congressional Election Survey, which has previously been linked to Catalist voting records and therefore allows us to compare block group income with self-reported income (Ansolabehere and Hersh, 2012). Figure A-4 shows that comparison. Apart from the close relationship between the income measures displayed in this graph, we also learn from the CCES that the block group measure picks up on other aspects of income class. For example, 75% of voters in block groups in which the median income is greater than \$100,000 report owning their home. In block groups in which the median income is less than \$20,000, only 35% report owning their home. Voters in poor block groups are also much more likely to report that jobs in their local area are scarce.

## **D OLS Regression Analysis of Income-Party Relationship**

To reinforce the intuition of the nonparametric averages, we present least squares regressions comparable to the nonparametric regression, using the full population of registered Democrats and registered Republicans. We first regress the binary dependent variable, Republican registration, on a registered voter's block group income and district income, employing state fixed-effects. The block group income variable takes on values 1, 3, 5, 7, 9, and 11, corresponding to the center of each of six major income

Figure A-4: Relationship between Self-Reported Family Income and Census Block Group Median Household Income



Means, 95% confidence intervals and interquartile ranges are shown. Data generated from 2008 CCES matched to Catalist Census data for 24,107 CCES respondents who were identified in registration records. The CCES asks respondents to report income within the ranges identified on the y-axis.

categories: \$0-\$19,999, \$20,000-\$39,999, and so on, up to a top category of \$100,000 and up. In a second regression, we add as a covariate the black proportion of district population. In a third regression, we add the interaction between the district income and district race measures.

The results of this analysis (Table A-1) convey the importance of district racial composition to the income-party relationship and show that the effect of district income evaporates once district race is accounted for. The first column shows that, after accounting for state fixed effects, Census block-group income, and state house district median income, Republican registration increases by 2.5 percentage points for every additional \$10,000 in median block-group income, and by an additional 1.7 percentage points for every additional \$10,000 in district income.

While the Model 2 and Model 3 coefficient on block-group level income remain identical to Model 1, the state-house-district income effect is reversed in Model 2 and reduces to 0 in model 3. This change has been picked up in the district-level race term. In model 2, Republican registration is now expected to drop by 6.3 percentage points for every additional 10 point increase in the black share of the population, a huge effect size. It is especially noteworthy that the addition of the contextual race variable does little to affect the coefficient on the block-group income variable, but erases the contextual income effect.

## **E Disaggregating Nonparametric Results by State**

While the district-level analysis presented in Figure 3 makes clear that race explains differences in the relationship between contextual income and partisan identification, it does so by disregarding the labels of the rich and poor states in which different types of districts are located. We now disaggregate the result by presenting summaries of the four sets of estimates that we presented in Figure 3, but representing them now by displaying the low-income and high-income estimates in each state, in each racial-composition interval. These results appear in Figure A-5. In each group, rather than presenting the average Republican registration in each \$20,000 income interval, we present for each state, the Republican registration in block groups with less than \$20,000 income and those with over \$100,000 average median block-group income. The Republican registration in low-income block groups in each type of state house district is represented by an “L,” and in high-income block groups it is represented by an “H.” The income category

Table A-1: Republican Party Identification by Race and Income

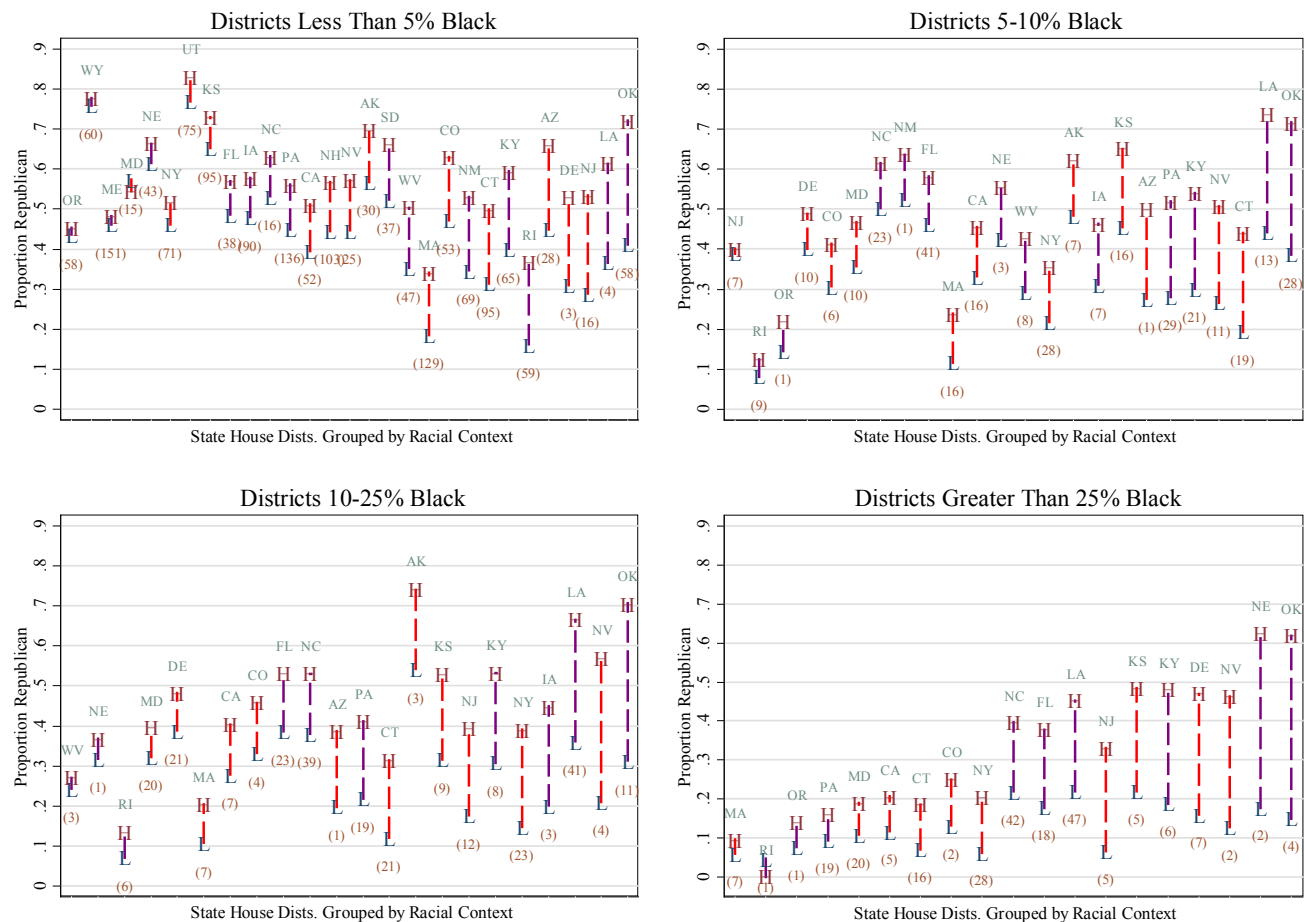
	(1)	(2)	(3)
	Dep. Var: Pr(Republican)		
Block-Group Income (in \$10K)	0.025	0.025	0.025
State House (SH) Median Income (in \$10K)	0.017	-0.004	0.000
SH Prop. Black		-0.633	-0.408
SH Income x SH Prop. Black			-0.062
Constant	0.209	0.378	0.362
Observations	73,170,970	73,170,970	73,170,970
R-squared	0.357	0.569	0.574

Note: All coefficients have standard errors less than 0.001. Coefficients on the explanatory variable of interest appear in red.

of each state is represented by red and purple. The number of districts in each state that meet the stated racial composition criteria appear in parentheses with each interval estimate. Notice that all 29 party registration states are represented in the upper left subplot, but some are missing from the other sub-plots since they lack districts that meet the racial criteria.

Presenting effect sizes separately by state confirms the intuition conveyed when the data from rich and poor districts are completely pooled. Voters in the racially homogeneous areas of rich and poor states are not substantially different in their voter registration behavior. In some states, the rich-poor partisan gap is bigger than others, but the gap is as big in New Jersey as Louisiana and as big in Connecticut as Kentucky. The states only sort into the familiar red-state/blue-state rich-state/poor-state categories with respect to income-based voting when we confine the analysis to districts containing large numbers of African Americans. In such districts, the gaps between the richest and poorest voters are smallest in states considered traditional “blue states”: Massachusetts, Rhode Island, Oregon, Pennsylvania, and Maryland. As we demonstrate with maps, the blacks who live in these blue states overwhelmingly live in

Figure A-5: State-Level Variation: State-Level Differences Rooted in Racially Heterogeneous Districts

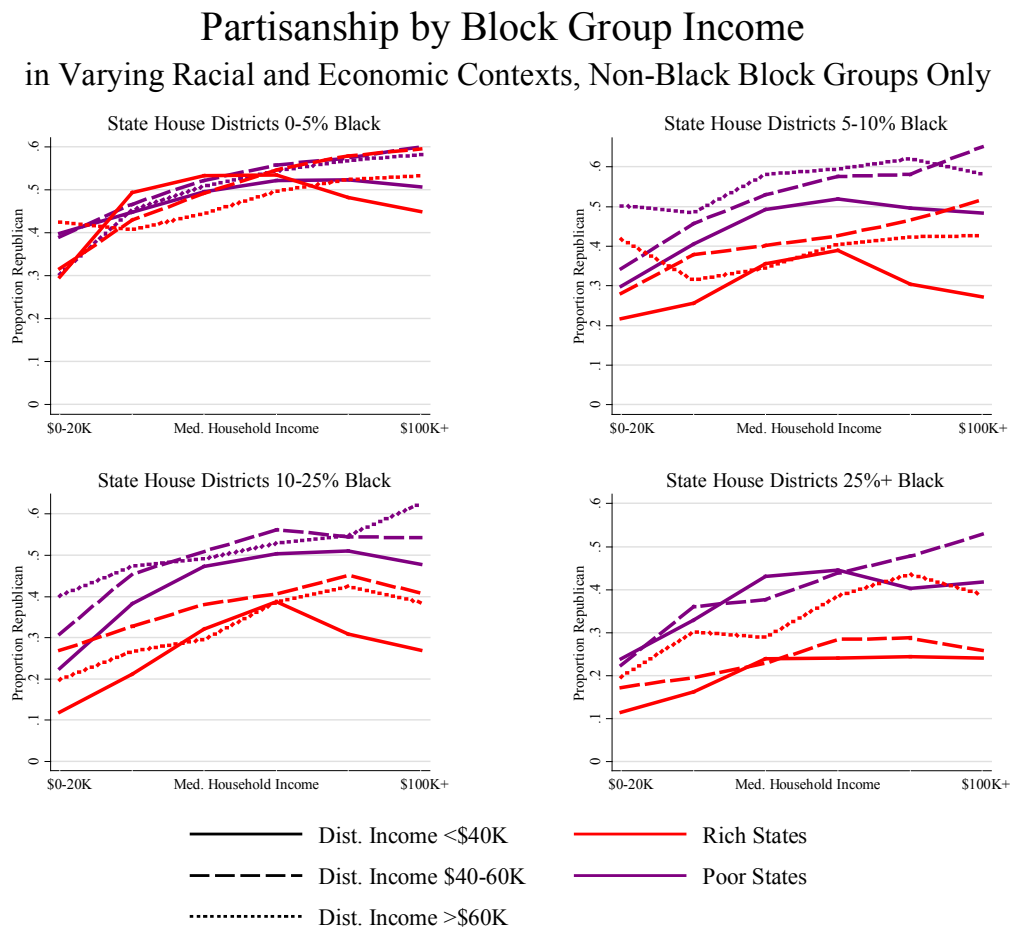


Note: “L” represents voters in low-income (\$0-20,000) block groups, while “H” includes voters in block groups with median income of \$100,000 and above. Purple lines are poor states, red lines are rich states. Number of districts included in calculation appears in parentheses. Figures are derived from the full population of registered Democrats and Republicans in party registration states. N= 73,170,970. Percentage black and median household income are derived from Year 2000 Census block group statistics; partisanship is derived from individual-level voter file records.

urbanized districts where non-black voters are also heavily Democratic. On the other end of the scale are a set of traditionally Republican states where blacks live in rural poverty and wealthy non-black voters living in proximity are much more Republican.

## F Replication of Figure 3 in for Non-Black Block Groups

Figure A-6: The Link Between Income Effects and Local Racial Context Persists When Only Non-Black Block Groups Included

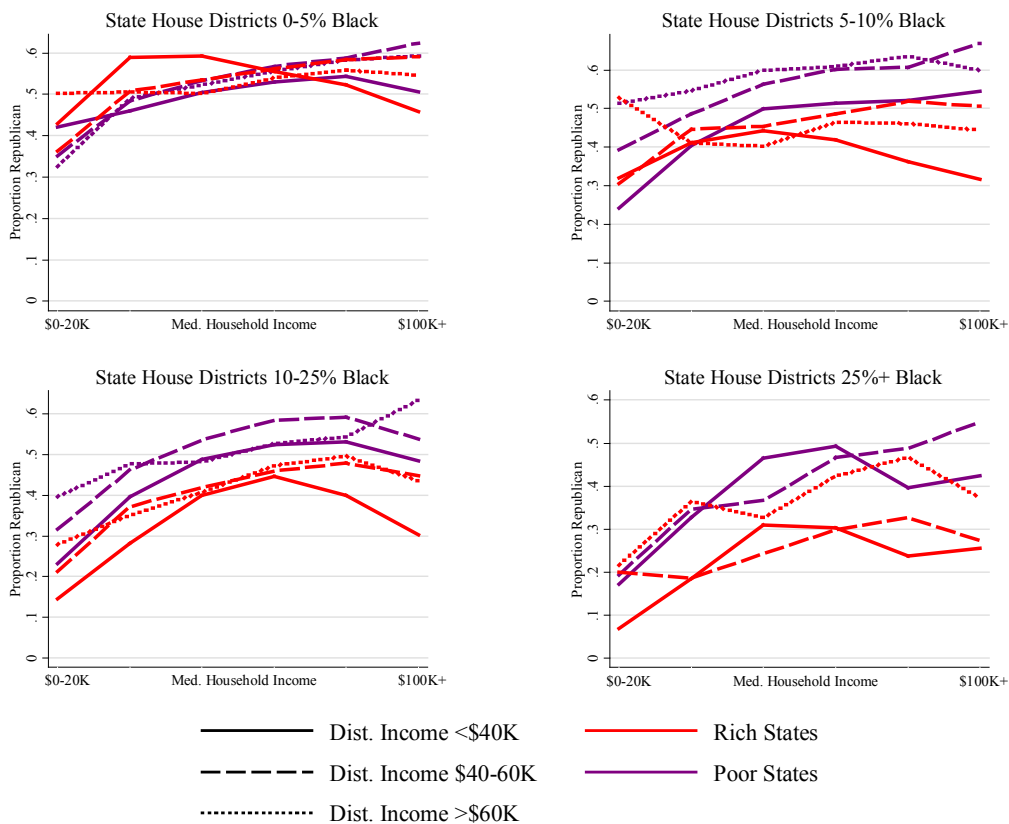


Note: Only registrants residing in block groups less than 10% black are included.



Figure A-7: The Link Between Income Effects and Local Racial Context Persists When Only White Non-Hispanic Block Groups Included

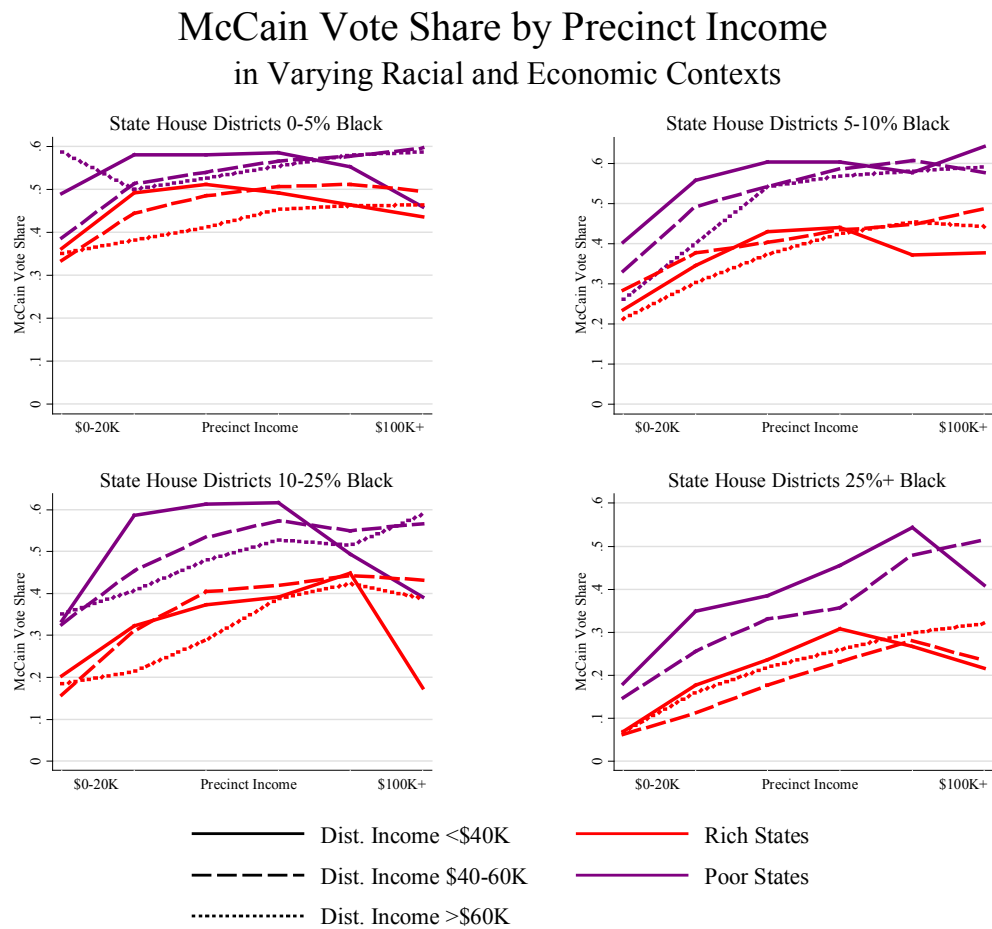
### Partisanship by Block Group Income in Varying Racial and Economic Contexts, White Non-Hispanic Block Groups Only



Note: Only registrants residing in block groups more than 90% non-Hispanic white are included.

## G Replication of Figure 3 with Precinct Returns

Figure A-8: The Link Between Income Effects and Local Racial Context Persists When Precinct Data are Used in Place of Registration Data



Note: This graph replicates the analysis if Figure 3 with precinct data from 49 states. For each precinct, we generate an income level based on the block group(s) located in that precinct. We show the relationship between precinct income and McCain vote share in the 2008 election. As in Figure 3, this relationship is shown in three levels of state house district income, four levels of state house district race, and two levels of state income. The same divergence between poor states and rich states emerges in the heavily black districts here as in the registration data.

## H Summary Statistics for Multi-Level Models

Summary statistics for the state house district data used in multilevel modeling appear in Table A-2.

Summary statistics for the HEDA data appear in Table A-3.

Variable	Mean	SD	Min.	Max.
Block Groups Per District	404	447	17	3,198
Two-Party Registrants	28,460	31,746	1,149	230,600
Proportion R Regis.	0.432	0.189	0.030	0.941
Proportion Black, 2000	0.080	0.146	0	0.928

Table A-2: Summary statistics for state house districts used in multilevel modeling (n=2,570).

Variable	Mean	SD	Min.	Max.
Total Votes	766	615	1	15,020
McCain Votes	351	364	0	8,419
Obama Votes	415	357	0	11,290
McCain Proportion of 2-Party Vote (%)	45	21	0	100
Median HH Inc., 1999 (x\$10,000) (Mean of BG Med HH Inc)	4.59	2.24	0	20
Pct. Black, 2000 (Total Black Pop./Total Pop.)	11.2	20	0	99.5

Table A-3: Summary statistics of Harvard Election Data Archive precinct data merged with 2000 Census block-group data (n=165,631).

## I Details on Multi-Level Modeling Using the lme4 Package

The multilevel model using the Catalist data translates into the following model in `lmer`:

```
lmer.out <- glmer(pctr2p~incgroup+(1+incgroup|stdist),  
weights=tot2p, data=shpanel)
```

where `pctr2p` is the Republican proportion of two-party registration, `incgroup` is a factor variable for income group, and `stdist` is a grouping variable for each state house district. Each of the district-income bin observations in the panel data is weighted using the variable `tot2p`, the total number of voters in each income category reporting affiliation with one of the two parties.

The multilevel model using the HEDA precinct data is represented via `lmer` as follows:

```
lmer.out<-lmer(pctr08~medhhinc10k+(1+medhhinc10k|county), data=precinct)
```

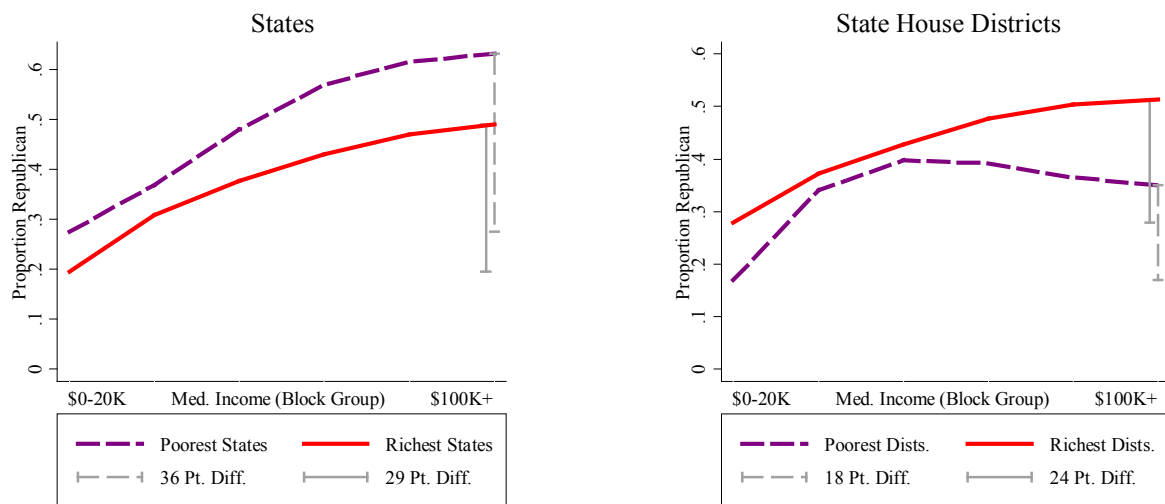
where `pctr08` is the McCain proportion of the two-party vote in each precinct, `medhhinc10k` is the unweighted average of the median household income in the block groups overlapping each district, measured in tens of thousands of dollars, and `county` groups precincts into counties.

### Effect of MAUP on the “Red-Blue Paradox”

To demonstrate how aggregation choices affect the analysis of income and partisanship, we conduct an exercise in which we replicate the key finding of Gelman et al. (2007) and then show that those findings do not map onto district-level analysis. In Figure A-9, we present results that compare states from the richest and poorest state-level income quartiles. The horizontal axis represents the block-group-level income category, while the vertical axis represents the proportion of two-party registrants registering as Republicans. In a second analysis, we aggregate the richest quartile of districts in all states and the poorest quartile of districts in all states. With these two groupings, we again measure the relationship between individual-level party registration and block-group income.

Results of this exercise confirm the findings in Gelman et al. (2008) and Gelman et al. (2007), but suggest that these results are driven by the use of states, rather than lower levels of aggregation, as a

Figure A-9: State House District Aggregates Follow Individual-Level Income-Party Patterns, State Aggregates Do Not



Note: Figures are derived from the full population of registered Democrats and Republicans in party registration states. N= 73,170,970. Median household income are derived from Year 2000 Census block-group statistics; partisanship is derived from individual-level voter file records.

measure of income context. The left panel of Figure A-9 presents Republican registration as a function of block-group income, stratifying according to state-level income. Echoing Gelman et al. (2008), the poorest voters (those in block groups with median income under \$20,000) register as Republicans at a rate of only twenty to thirty percent, with poorer voters in the poorer states actually slightly more likely to register Republican (8 percentage points more likely) than poor voters in richer states. However, as block-group level income increases, the gap between voters in rich and poor states grows. At the highest income levels, the gap in Republican registration and voting between rich states and poor states reaches 14 percentage points. Only 49% of members of the top income category (block-group income of \$100,000 or more) in rich states are registered as Republicans, while over 63% of the richest voters in poor states are registered as Republicans.

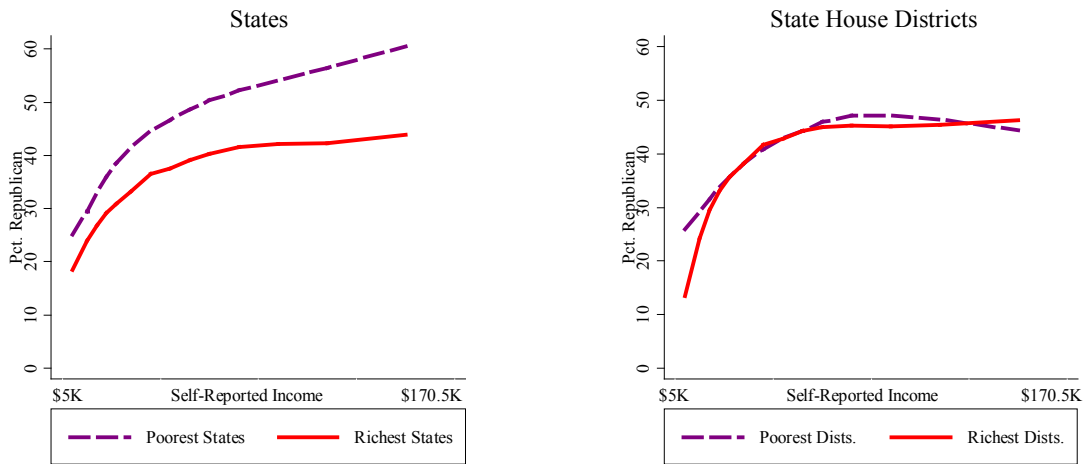
When we focus on state house districts as income contexts rather than states, the familiar “paradoxical” relationship between income context and income-based voting behavior flips. The right panel of Figure A-9 presents this result. The poorest voters in poor state house districts register as Republicans at a rate of only 17 percent while poor voters in richer state house districts register at an average rate of almost 28 percent. In rich state house districts, 51% of voters in the \$100,000-plus income bracket are registered Republicans, but only 35% of the richest voters in the poorest state house districts are registered Republicans, a number that is almost as low as the percentage of Republicans among poor voters in the richest districts.<sup>8</sup> Thus, there is something about the lower-level contexts in all states that sets them apart from states, and we would not expect predictions about the income-party relationship obtained using states to transfer to the lower geographic levels more typically used in scholarship on geographic context.

We can also replicate this result with survey data rather than party registration data. We replicate Figure A-9 using party identification and self-reported income rather than party registration and block group income. In making A-10, we divide the sample into state-level and district-level income quartiles. For these quartiles, we use block group income. For the richest and poorest states and the richest and poor-

---

<sup>8</sup>The thresholds for binning states or state house districts into rich and poor states are based on top and bottom quartiles of the average of the block-group median income in party registration states.

Figure A-10: Replication of Figure A-9 using Self-Reported Partisanship and Income Fields



Note: N= 10,095 self-reported weighted Democrats and Republicans in party registration states. Lines are created with lowess smoothing. Data comes from the 2008 CCES matched to census block via Catalist (Ansolabehere and Hersh, 2012).

est districts, we then measure the relationship between Republican self-identification and self-reported income. As in Figure A-9, the gap between rich and poor is larger in poor states than rich states but is bigger in rich districts than poor districts.

## J Variation in Precinct-Level Income-Based Voting Within Counties

The relationship between local racial composition and income-based voting becomes more obvious in the map of precinct-level income effects that are allowed to vary by county (Figure A-11). These results throw into sharper relief the heterogeneity that exists within different racial contexts. In key urbanized areas of the South, notably in areas such as Atlanta and Miami, the relationship between income and partisanship is relatively flat, while in areas along the Mississippi River the relationship is steep. This is of vital importance to claims about income-based voting within states. Were we to assemble a new state out of the metropolitan areas of the New South such as Atlanta and Nashville, the income-based voting

in these states would look much like states of the Northeast. If we instead constructed a state consisting of the Black Belt counties of North Carolina, Georgia, Alabama, Tennessee, and Florida, income-based voting would be much stronger in such states than in the states given to us by history.

The map of 2008 precinct-level results presents a few surprising results that differ from party registration figures. Some of these differences may be an artifact of the 2008 presidential election, but others suggest explanations related to the different measures. One is that the income-based differences in partisan registration that appear in registration data in California's Central Valley do not appear on this map. This may itself be an artifact of the modifiable areal unit problem (especially since counties in the Western region are typically larger than state house districts), but it may also relate to a partisan shift towards the Democrats in such counties due to the disproportionate impact of the 2008 foreclosure crisis. The other interesting finding is that across the rural and small-city areas of Pennsylvania, income-based voting is especially pronounced. Many of the Appalachian counties where the marginal effect of income on partisanship was larger than expected are, for example, counties in which Obama lagged Hillary Clinton in the primaries (Gibson and Gleason, 2012). But most of these deviations pale in comparison to the major polarization that appears on the map: voters in precincts in counties in the Northeast Corridor from Northern Virginia to New Hampshire engage in much less income-based voting than we would expect from precinct-level income alone, while the rural counties of the South and across isolated patches of the country's midsection engage in much more income-based voting. Our analysis suggests that the reason for this difference has to do with voting patterns in areas with blacks, not with areas that are homogeneously white.

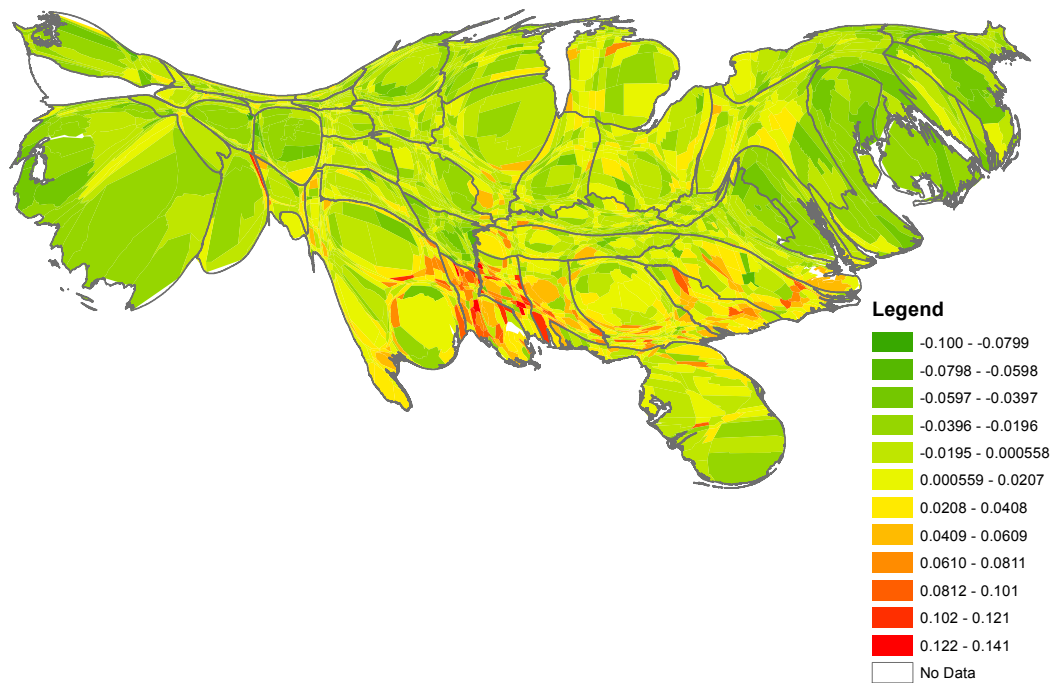
## **K State-by-State Analysis of Party Registration, Block-Group Income, District Income, and District Percent Black**

We provide the equivalent of Figure 3 for each individual state. The graphs appear at the end of the appendix. In these graphs, we are displaying conditional means analogous to coefficients from a regression model in which party affiliation is regressed on block group income (6 categories), district income



Figure A-11: Cartogram of Random Effects for Income, Precincts Within Counties

Within-County Random Effect of Precinct Income (Tens of Thousands of Dollars) on McCain Vote  
All Counties

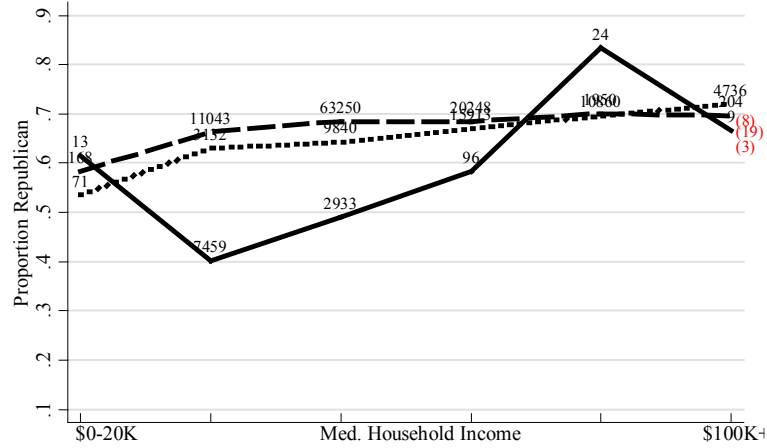


Cartogram of county-specific marginal effect of each \$10,000 of average block group median household income in each precinct on the McCain vote share in the 2008 presidential election. Varying shades of green indicate districts in which the correlation between block group income and partisan registration are below the national median among state house districts, while varying shades of red indicate places where this relationship is above the national median.

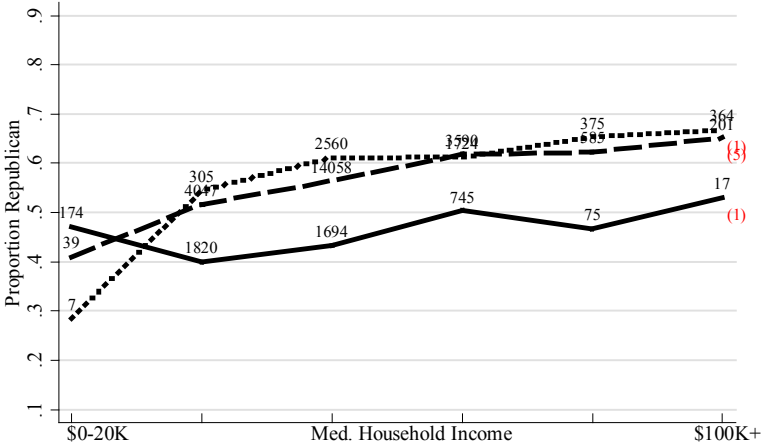
(3 categories), district race (4 categories), and state of residence (29 categories), with all independent variables interacted. Though some of these interactions have missing values, the graphs subdivide the 73 million-person sample into 2,088 subpopulations. In each graph that follows, the number of districts represented in each line is indicated in parentheses in red. The number of observations (Democratic and Republican registrants) in each cell is indicated above each estimate.

# Alaska

State House Districts 0-5% Black



State House Districts 5-10% Black

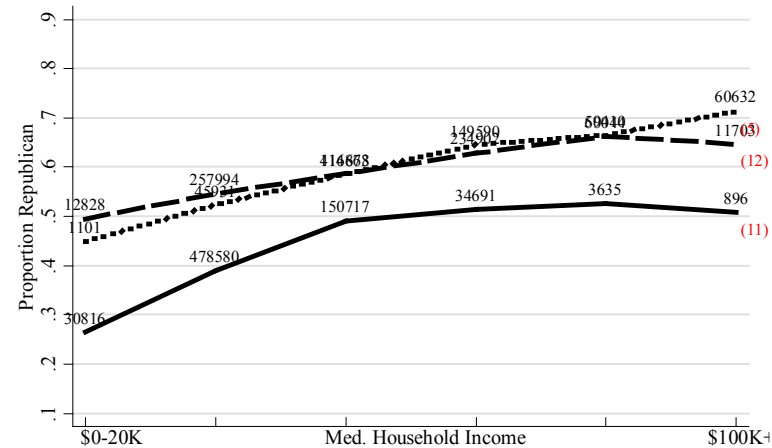


State House Districts 10-25% Black

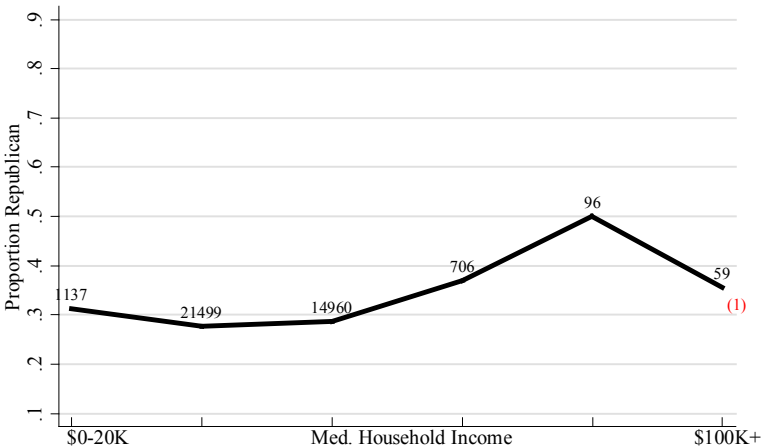


# Arizona

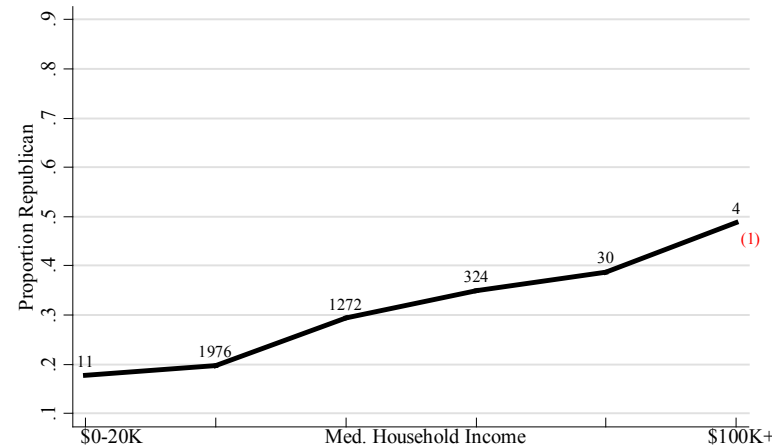
State House Districts 0-5% Black



State House Districts 5-10% Black

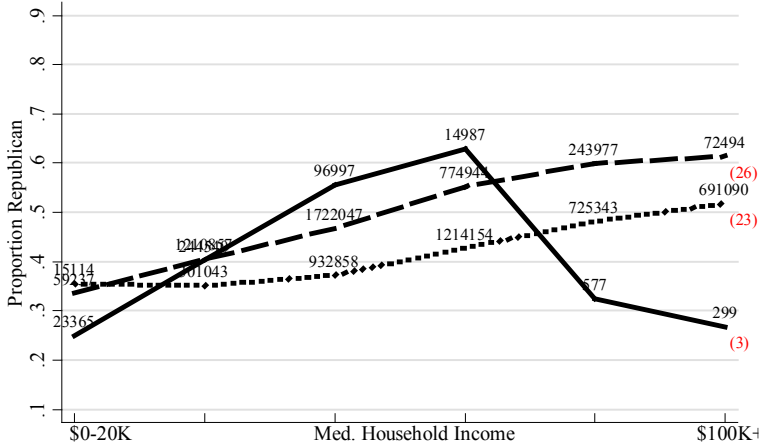


State House Districts 10-25% Black

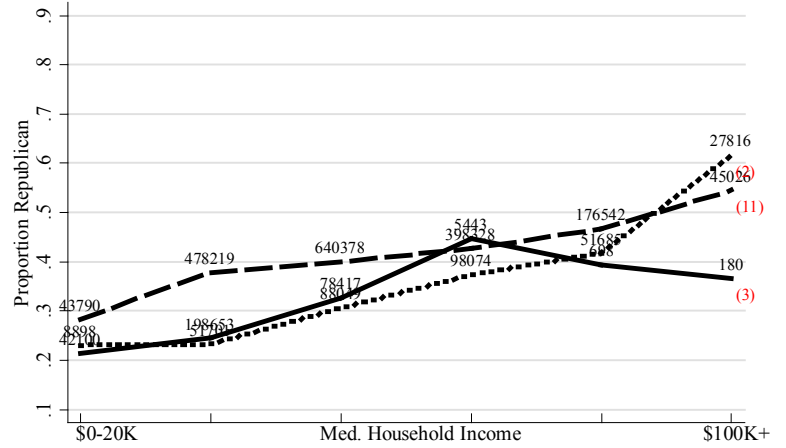


# California

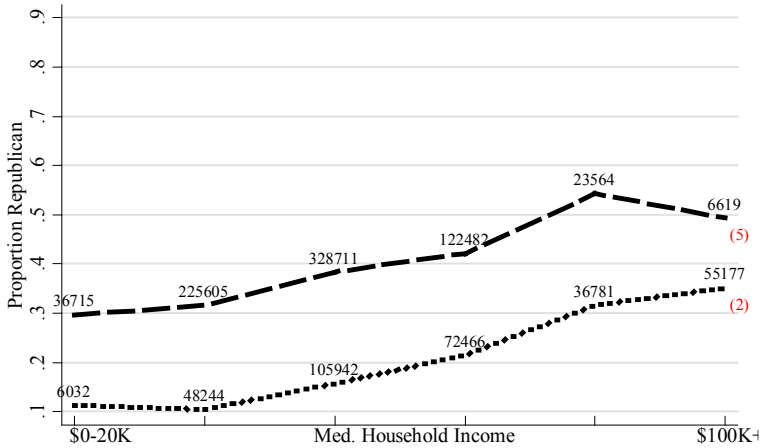
## State House Districts 0-5% Black



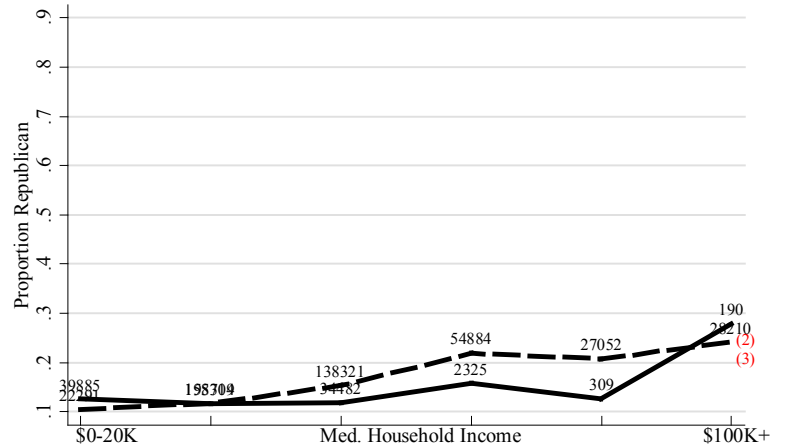
## State House Districts 5-10% Black



## State House Districts 10-25% Black

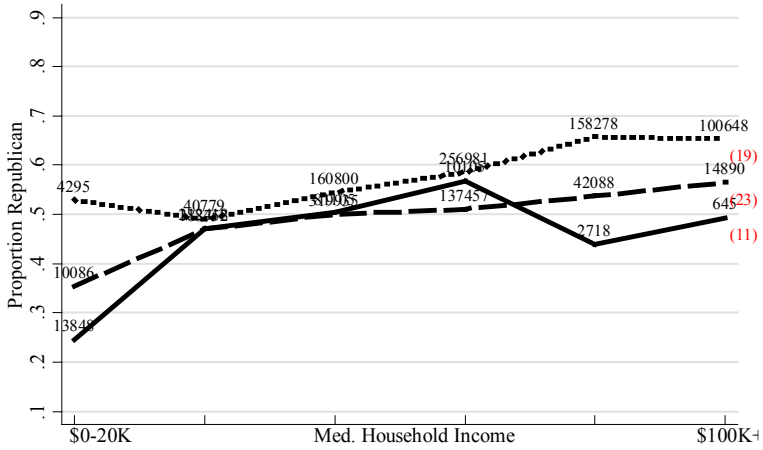


## State House Districts 25+% Black

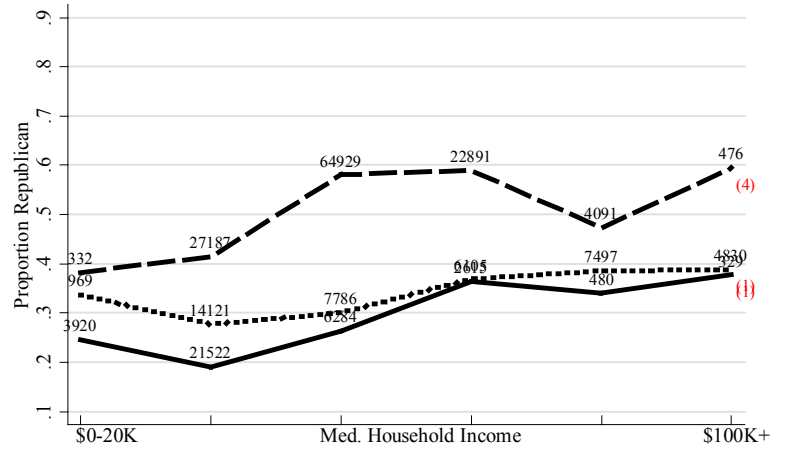


# Colorado

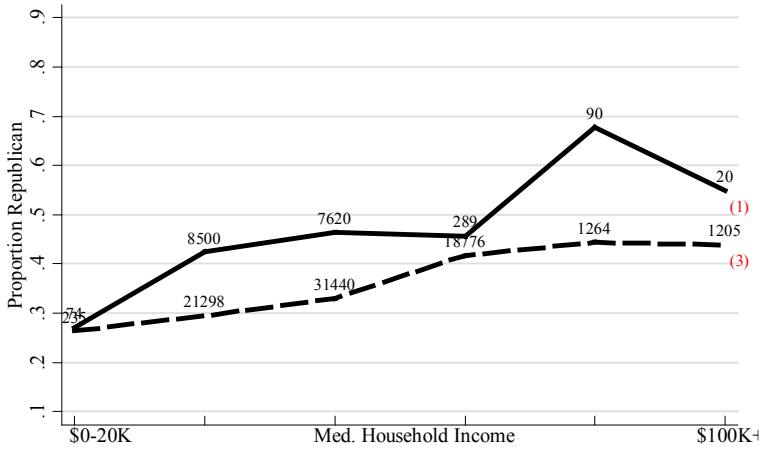
## State House Districts 0-5% Black



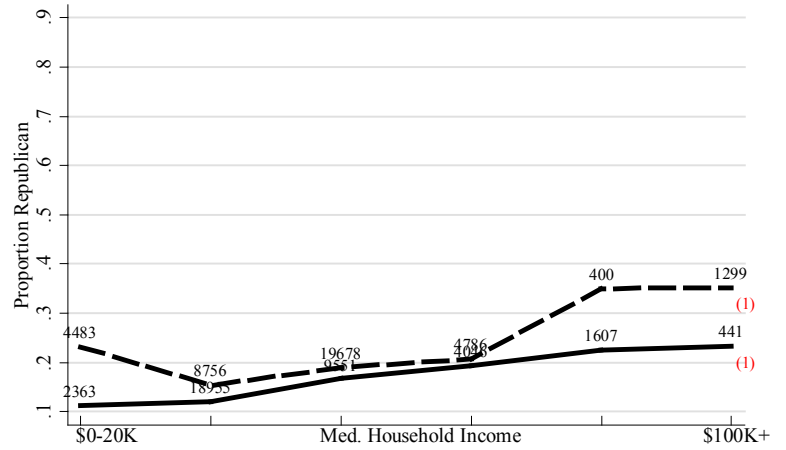
## State House Districts 5-10% Black



## State House Districts 10-25% Black

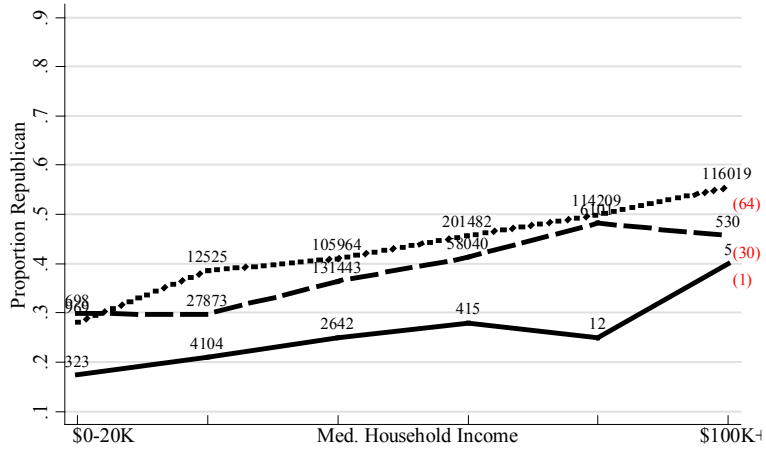


## State House Districts 25+% Black

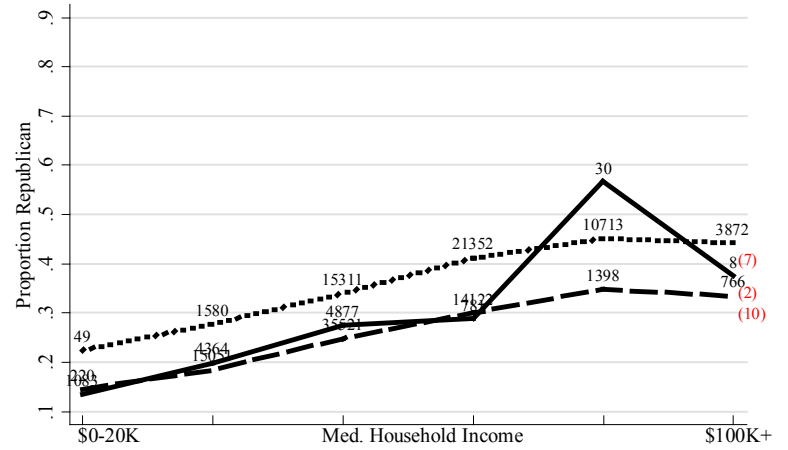


# Connecticut

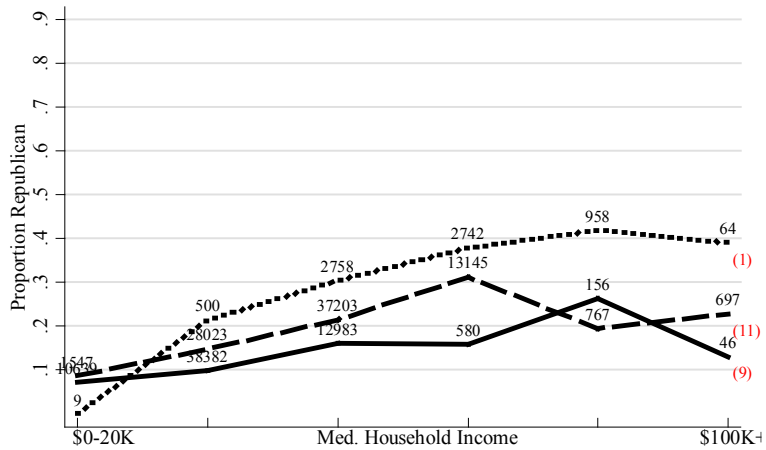
## State House Districts 0-5% Black



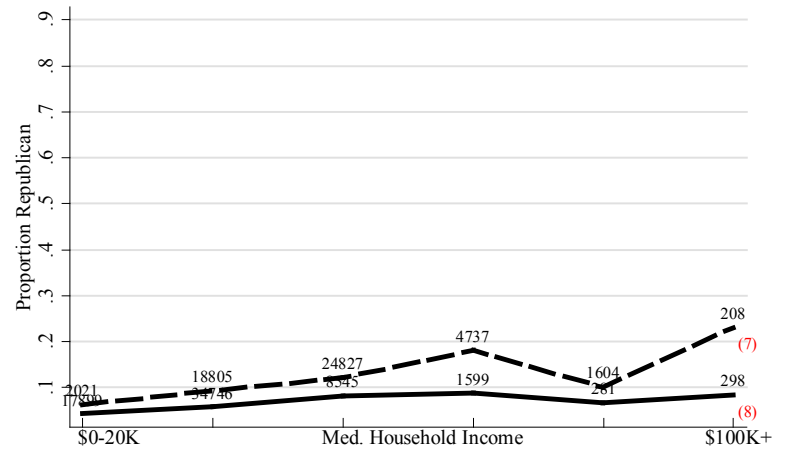
## State House Districts 5-10% Black



## State House Districts 10-25% Black

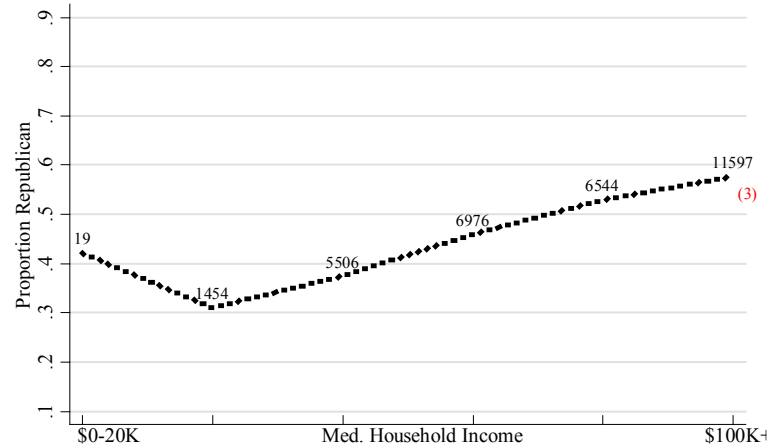


## State House Districts 25+% Black

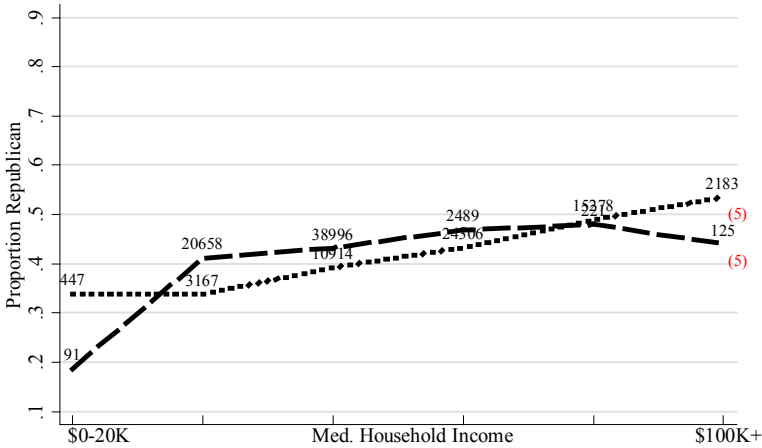


# Delaware

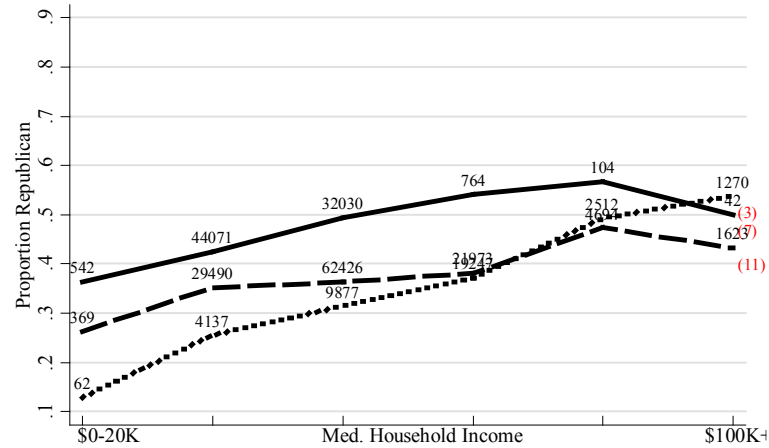
State House Districts 0-5% Black



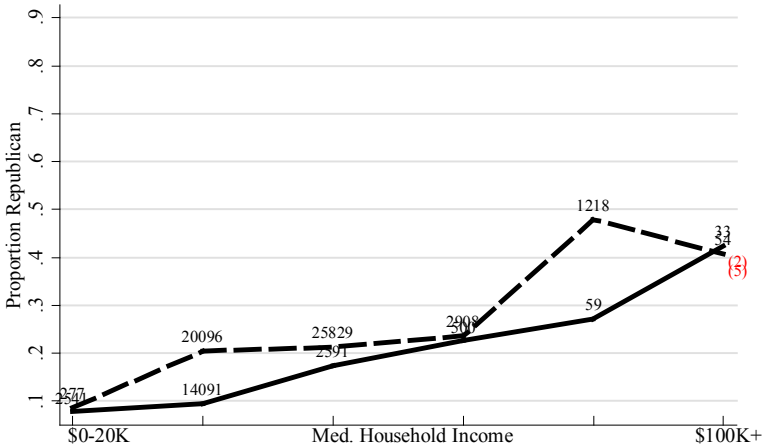
State House Districts 5-10% Black



State House Districts 10-25% Black



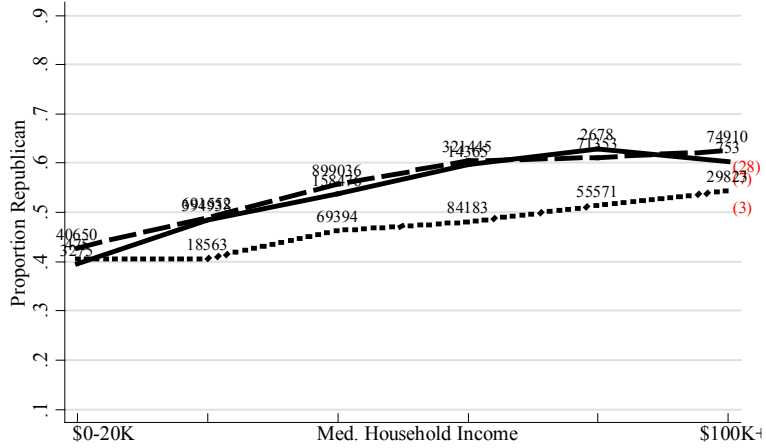
State House Districts 25+% Black



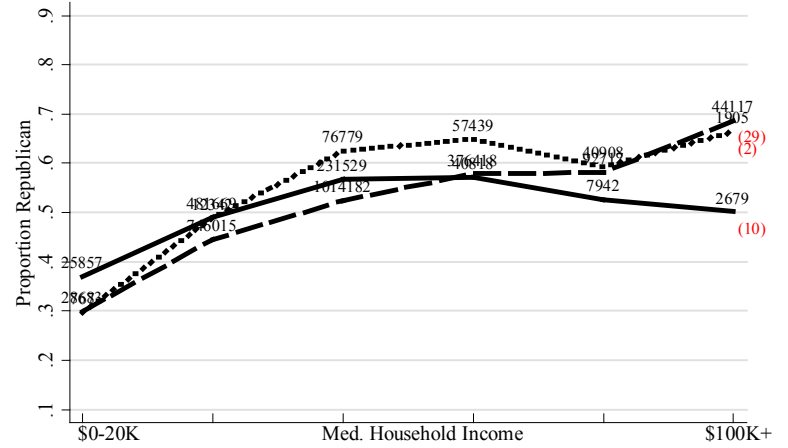


# Florida

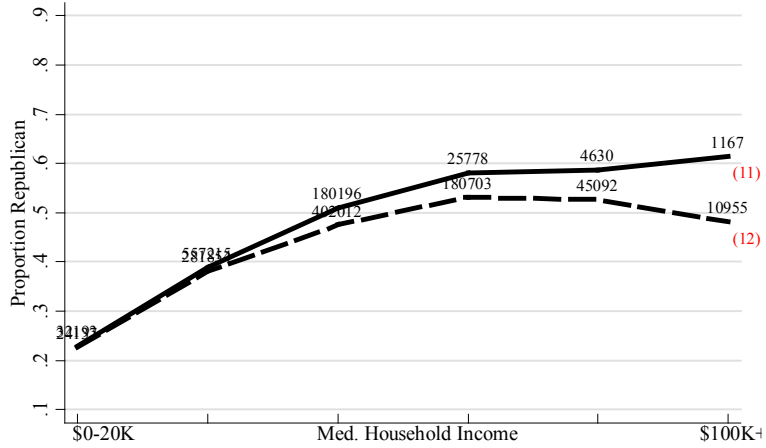
## State House Districts 0-5% Black



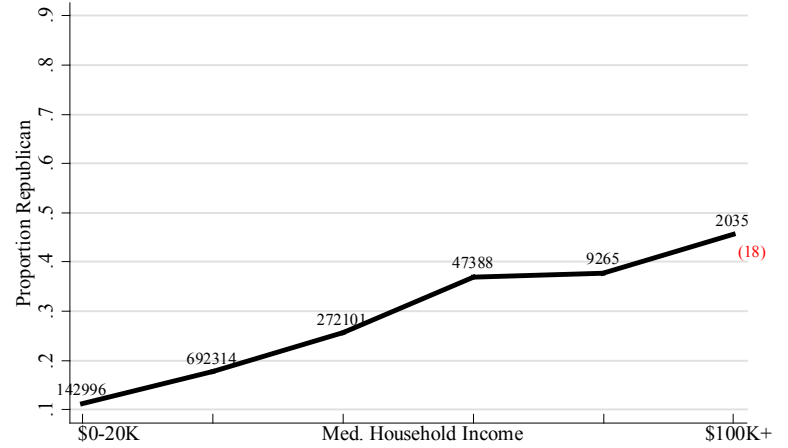
## State House Districts 5-10% Black



## State House Districts 10-25% Black

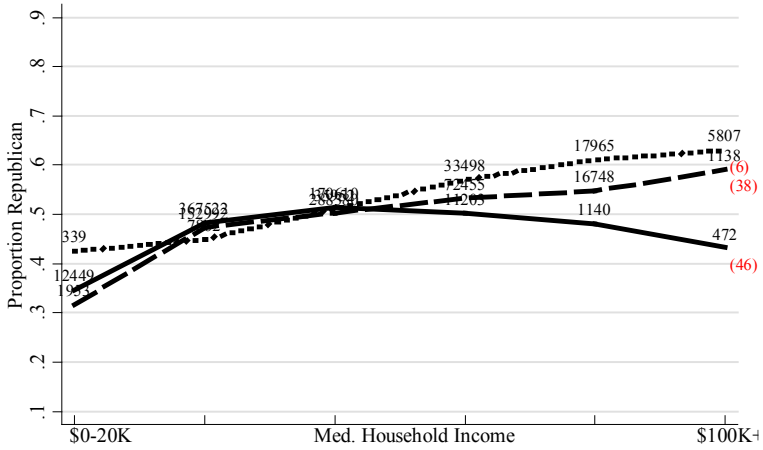


## State House Districts 25+% Black

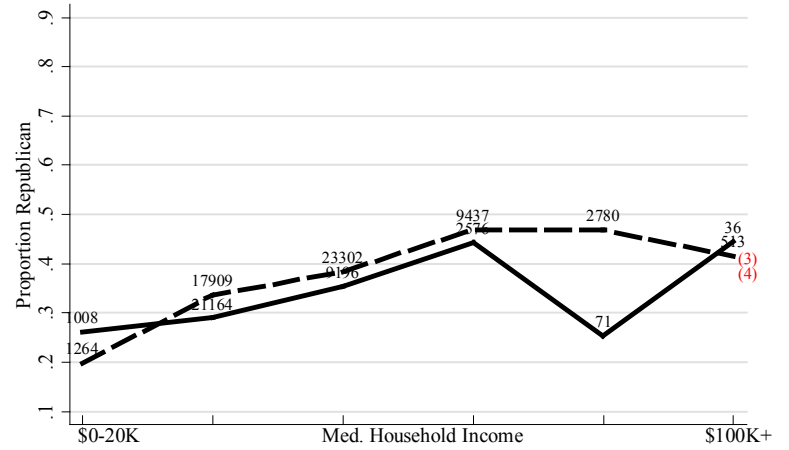


# Iowa

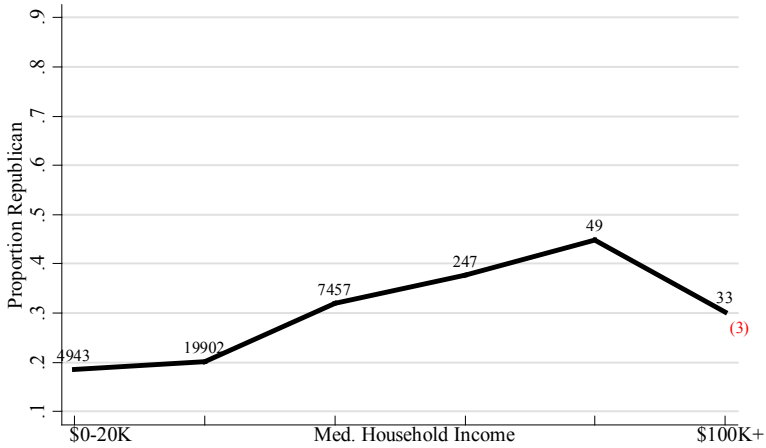
## State House Districts 0-5% Black



## State House Districts 5-10% Black

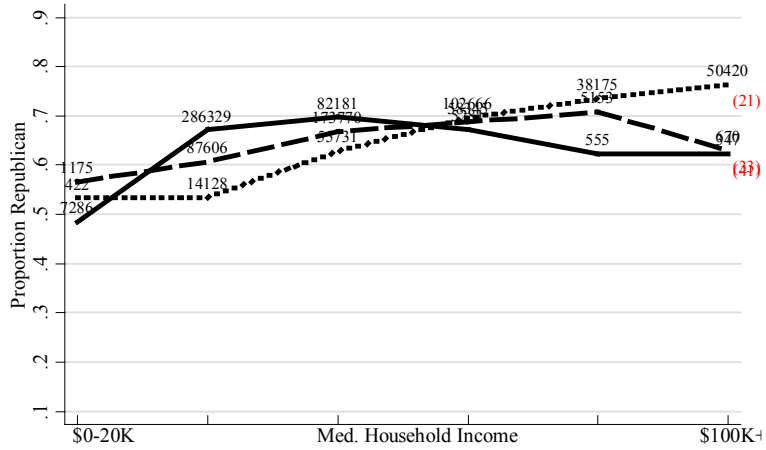


## State House Districts 10-25% Black

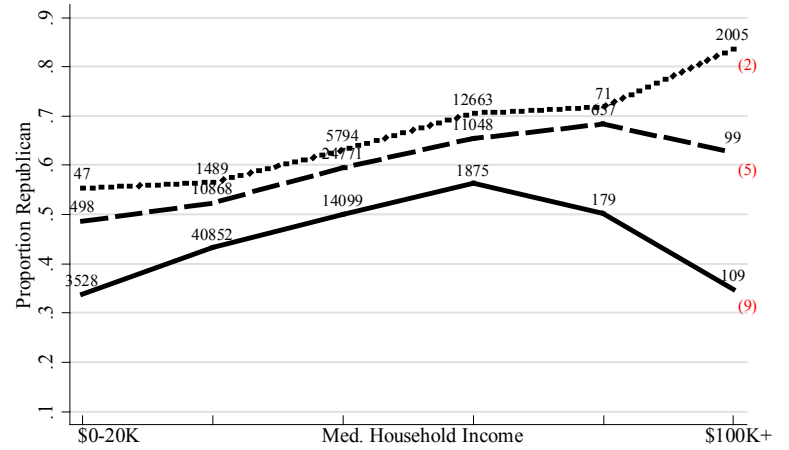


# Kansas

## State House Districts 0-5% Black



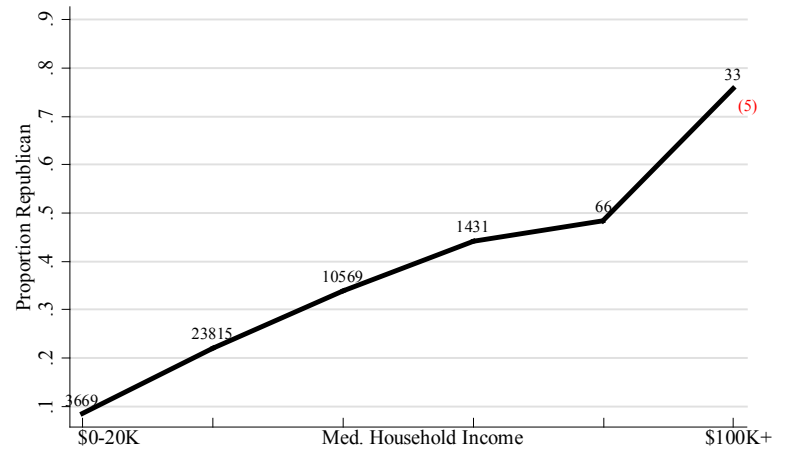
## State House Districts 5-10% Black



## State House Districts 10-25% Black

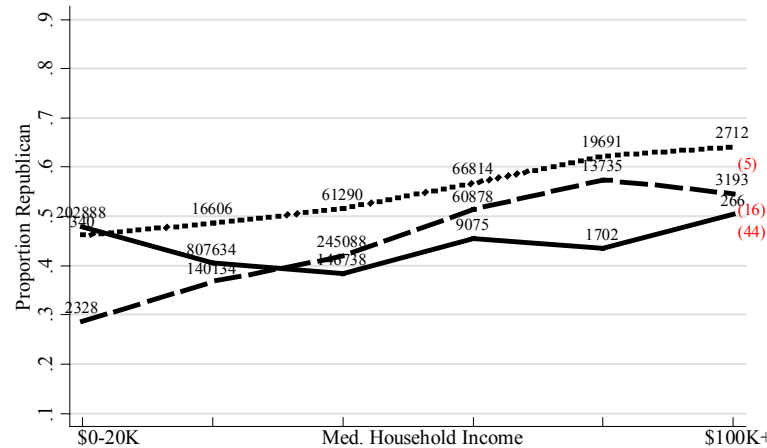


## State House Districts 25+% Black

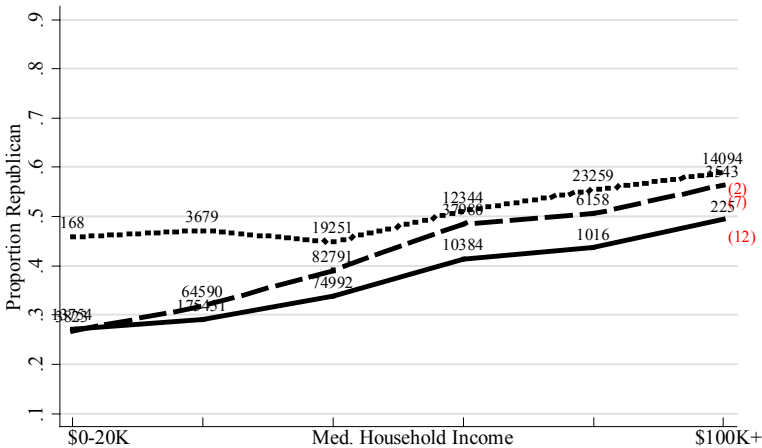


# Kentucky

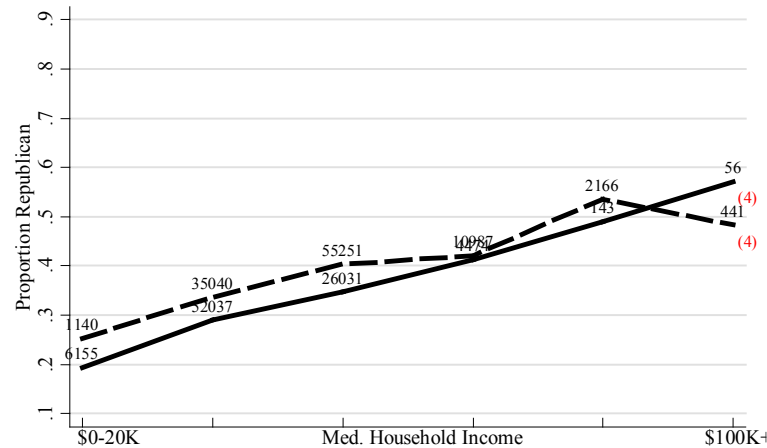
State House Districts 0-5% Black



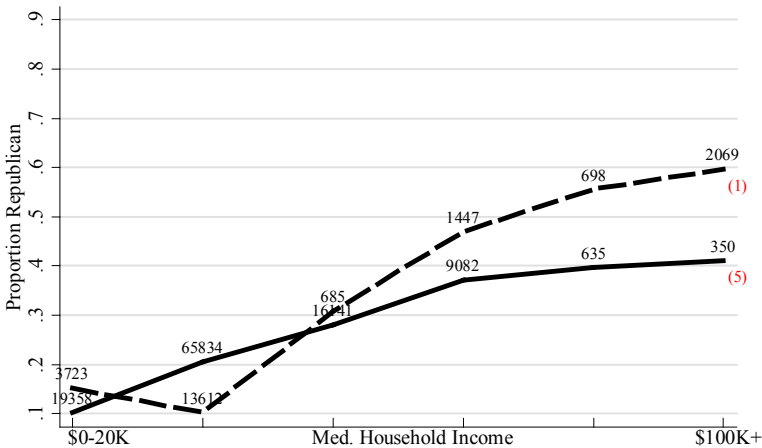
State House Districts 5-10% Black



State House Districts 10-25% Black

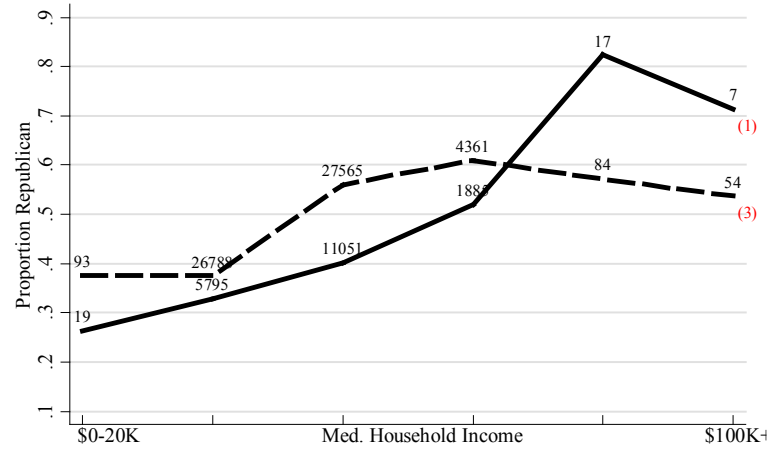


State House Districts 25+% Black

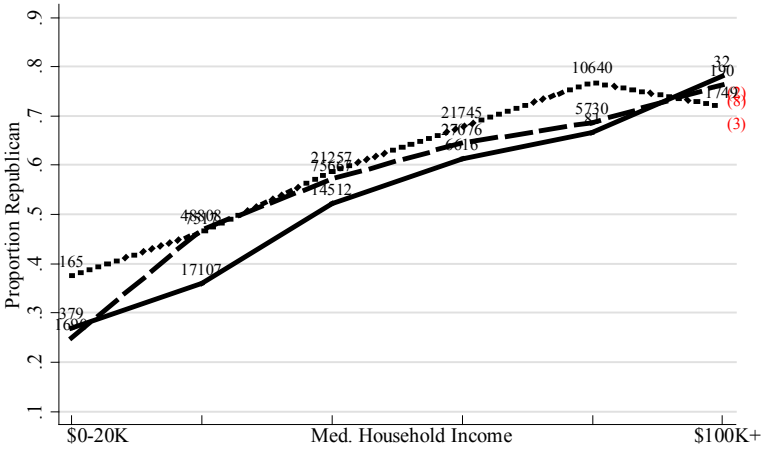


# Louisiana

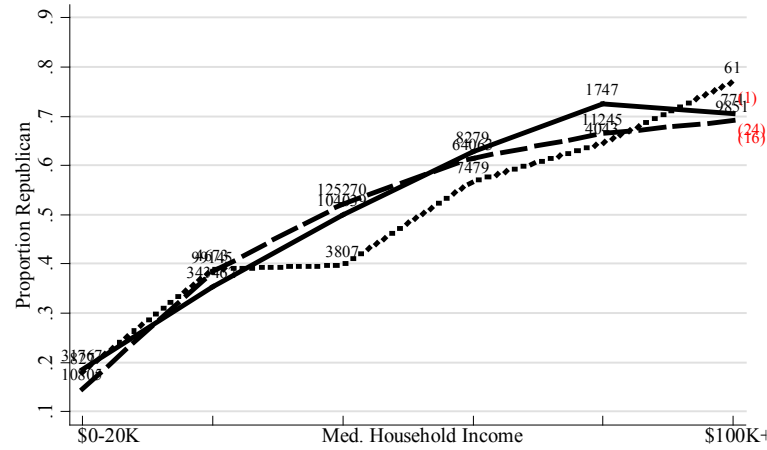
State House Districts 0-5% Black



State House Districts 5-10% Black



State House Districts 10-25% Black



State House Districts 25+% Black

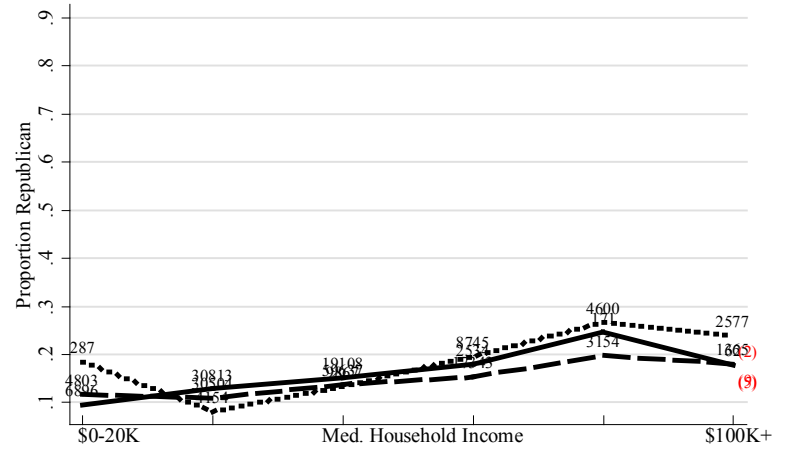


# Massachusetts

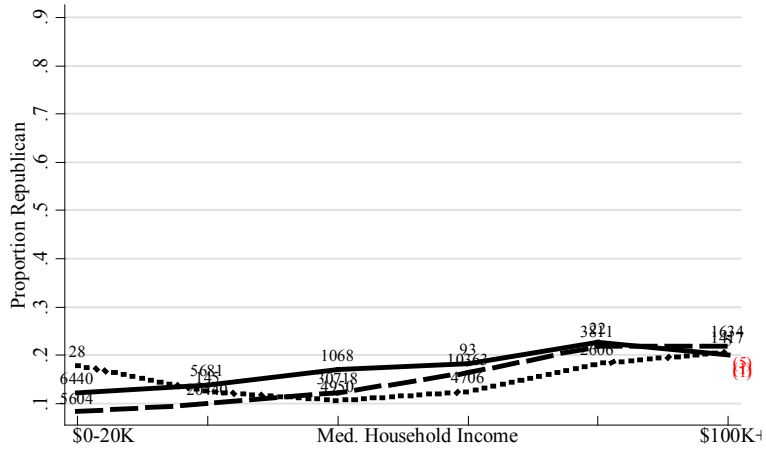
State House Districts 0-5% Black



State House Districts 5-10% Black



State House Districts 10-25% Black

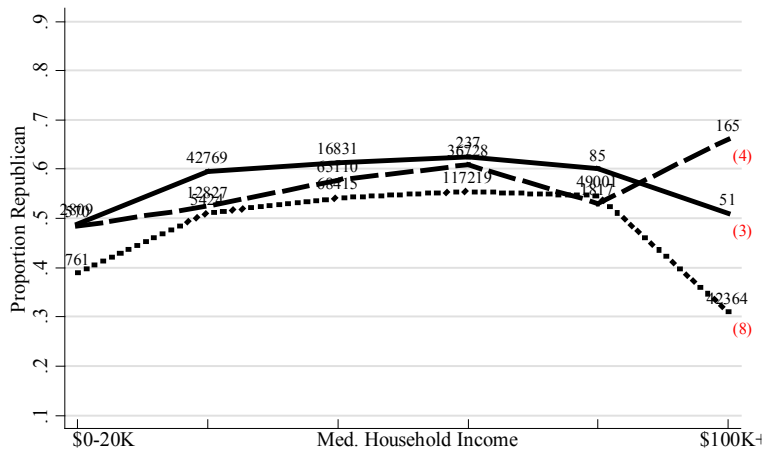


State House Districts 25+% Black

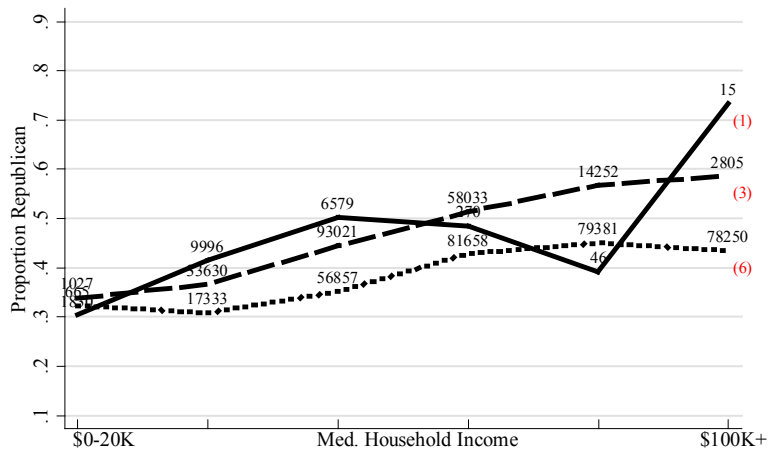


# Maryland

## State House Districts 0-5% Black



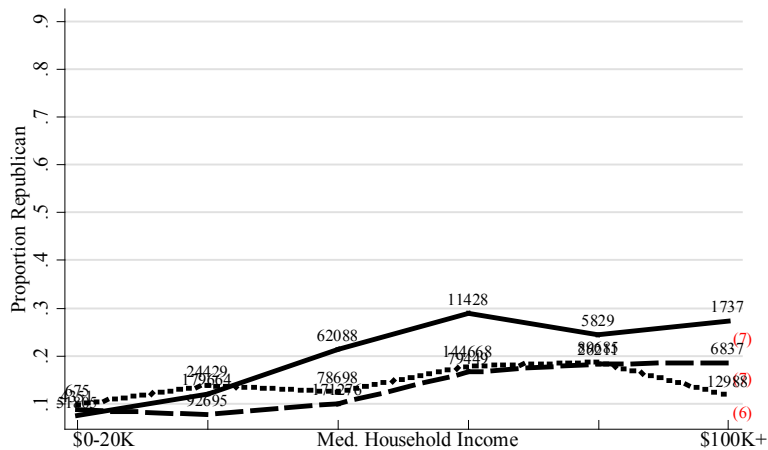
## State House Districts 5-10% Black



## State House Districts 10-25% Black

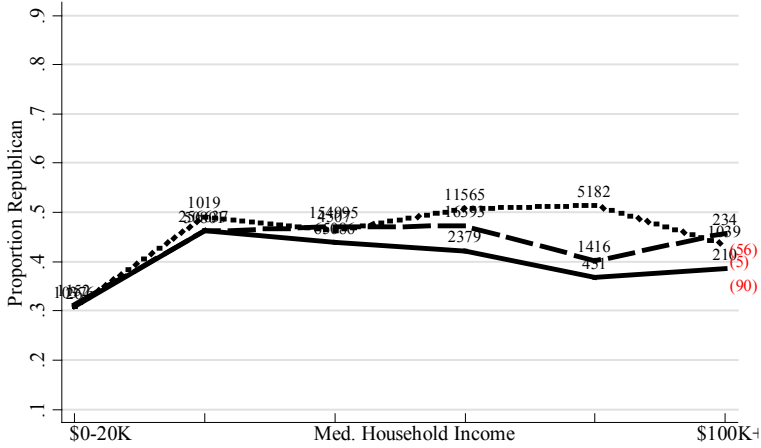


State House Districts 25+% Black

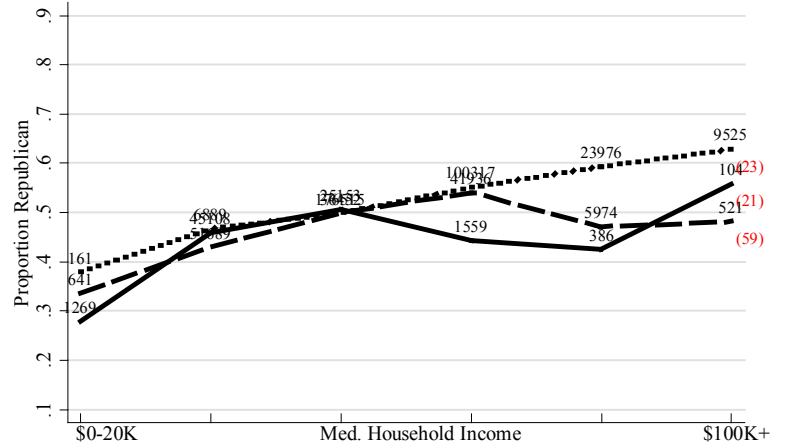


# Maine, New Hampshire, and New Mexico

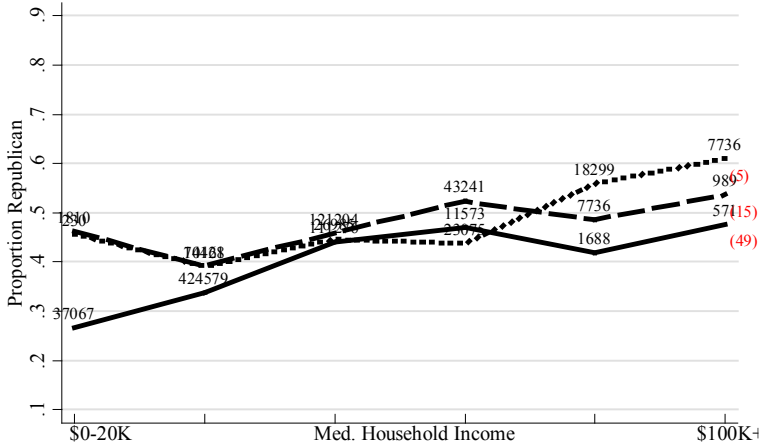
MAINE - State House Districts 0-5% Black



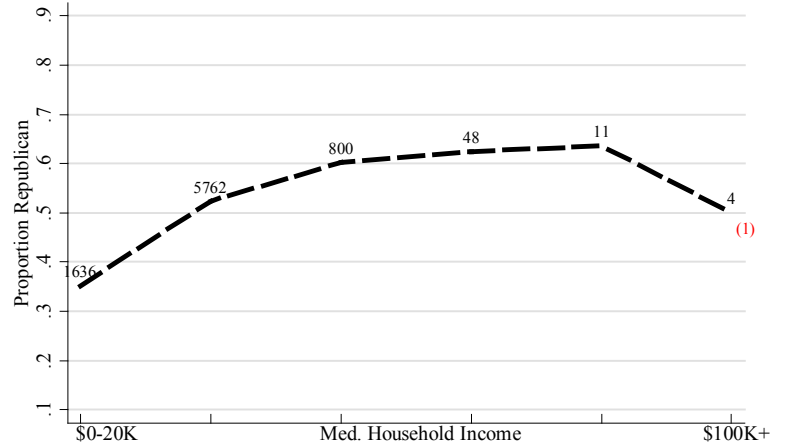
NEW HAMPSHIRE - State House Districts 0-5% Black



NEW MEXICO - State House Districts 0-5% Black



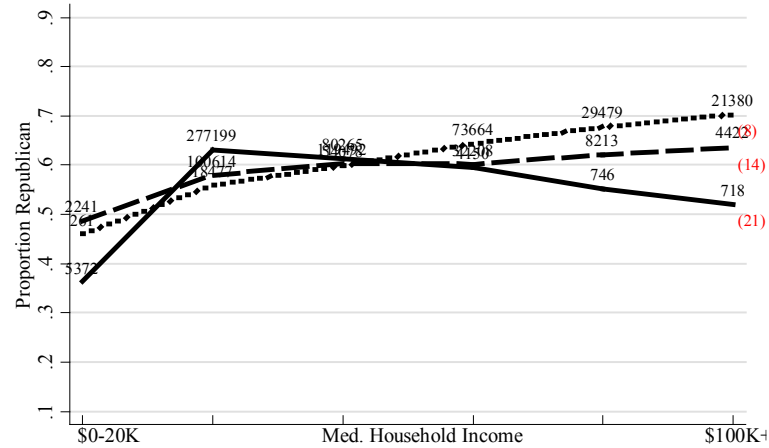
NEW MEXICO - State House Districts 5-10% Black



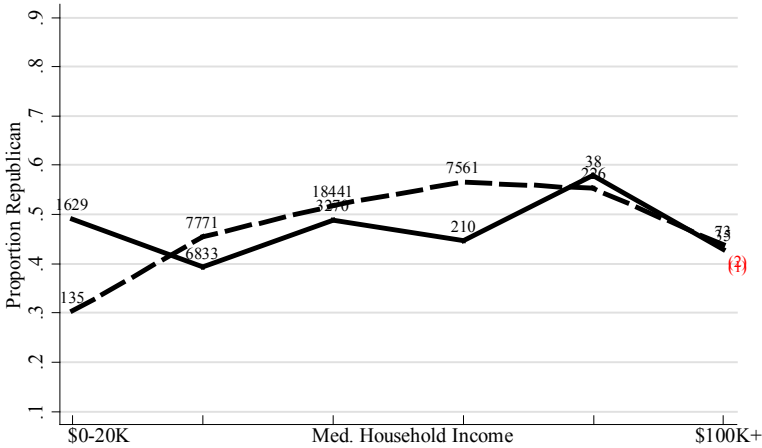


# Nebraska

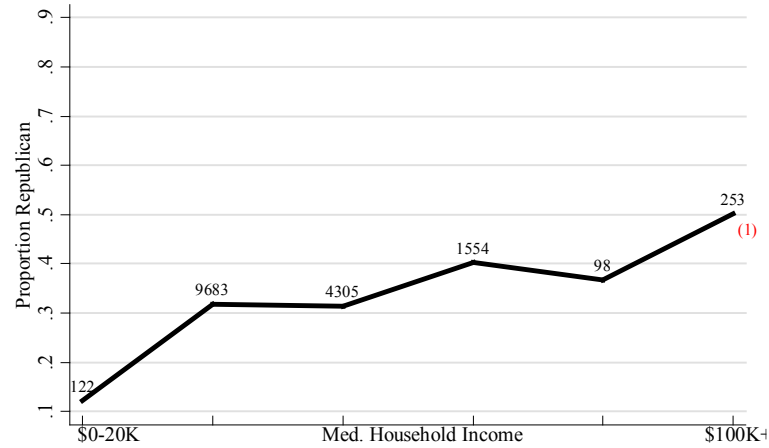
State House Districts 0-5% Black



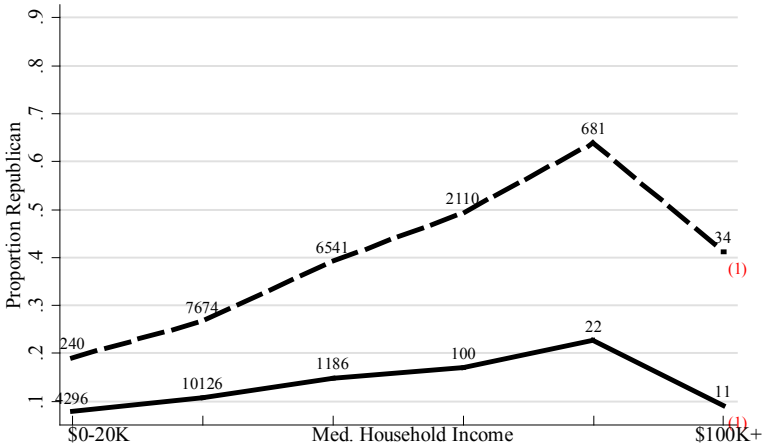
State House Districts 5-10% Black



State House Districts 10-25% Black

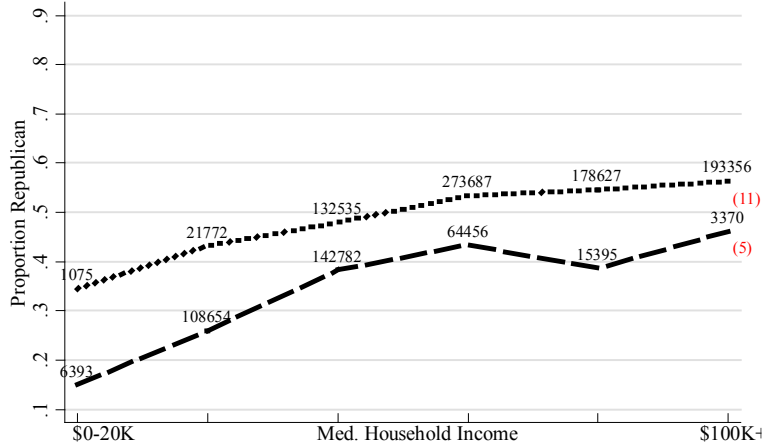


State House Districts 25+% Black

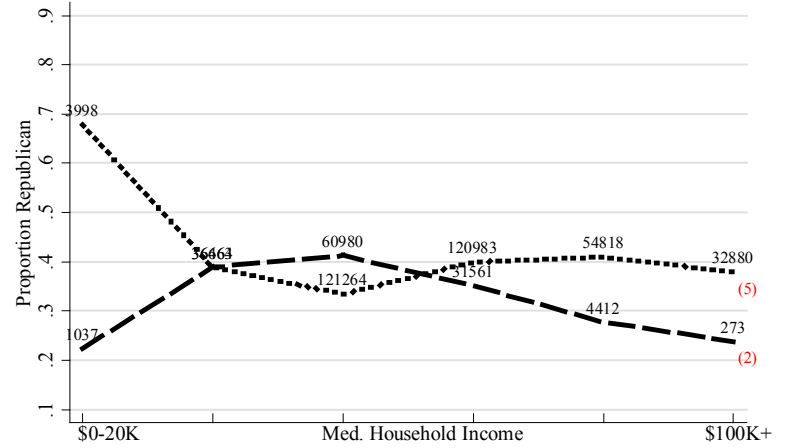


# New Jersey

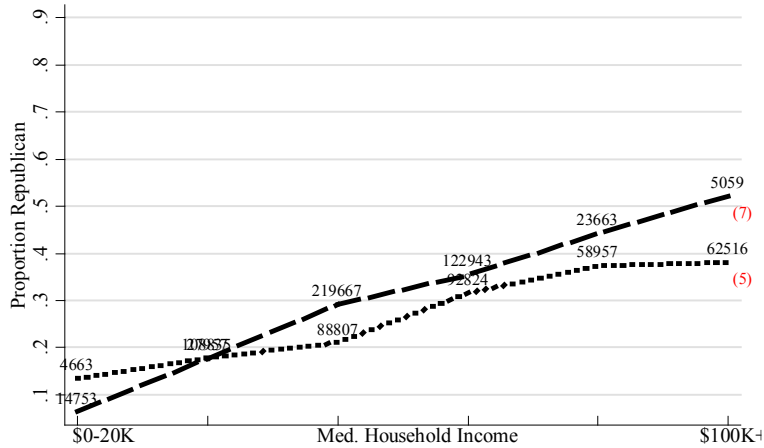
## State House Districts 0-5% Black



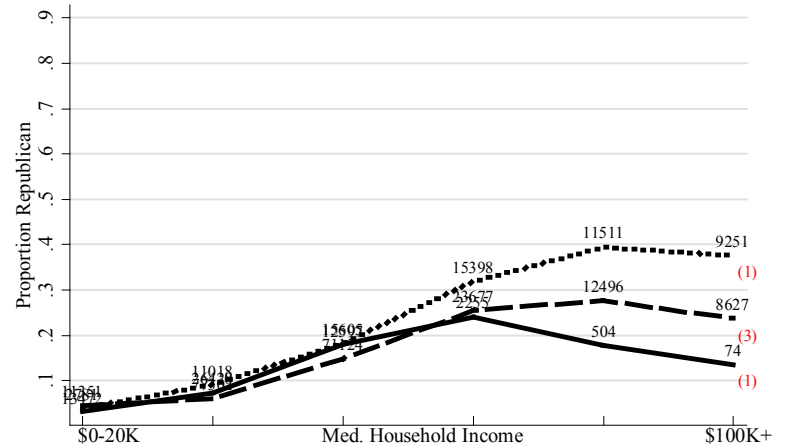
## State House Districts 5-10% Black



## State House Districts 10-25% Black

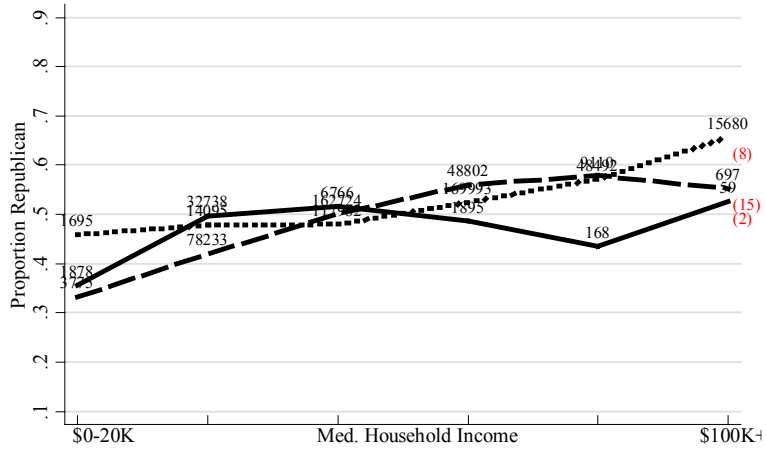


## State House Districts 25+% Black

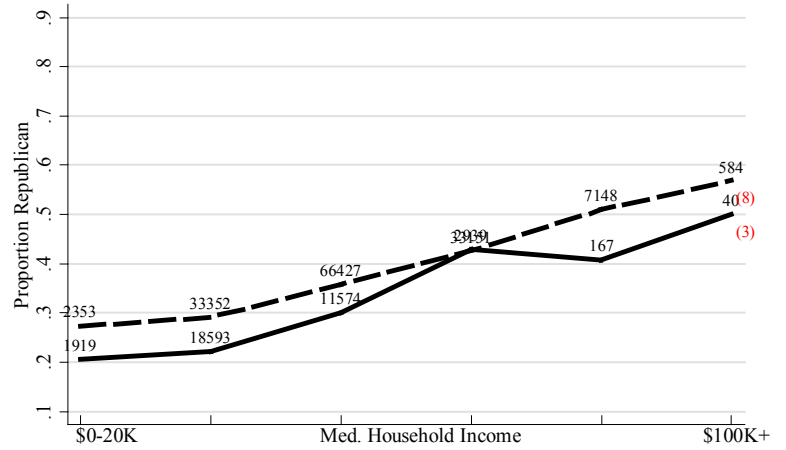


# Nevada

## State House Districts 0-5% Black



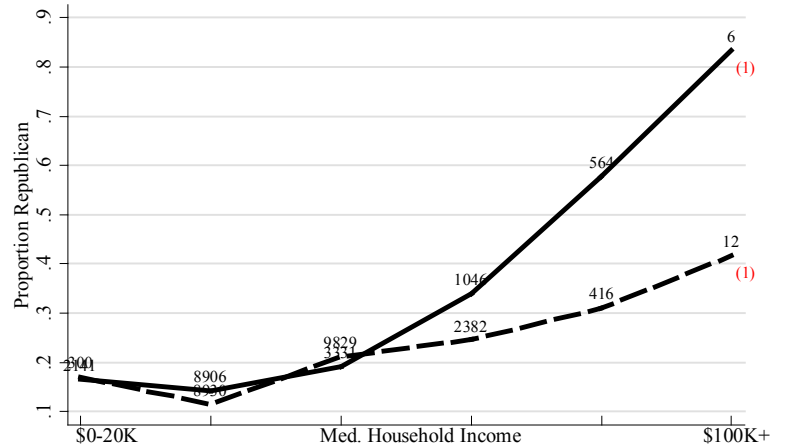
## State House Districts 5-10% Black



## State House Districts 10-25% Black

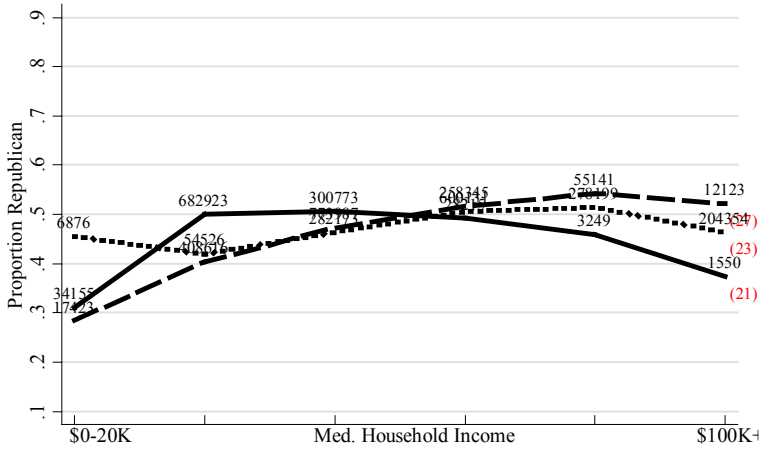


## State House Districts 25+% Black

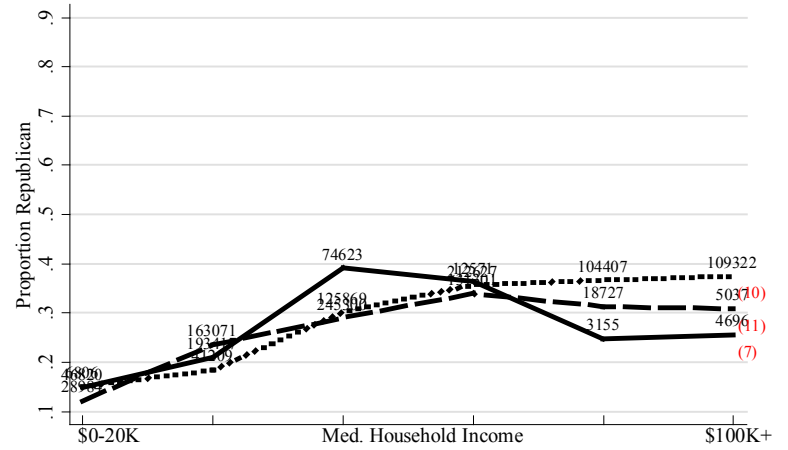


# New York

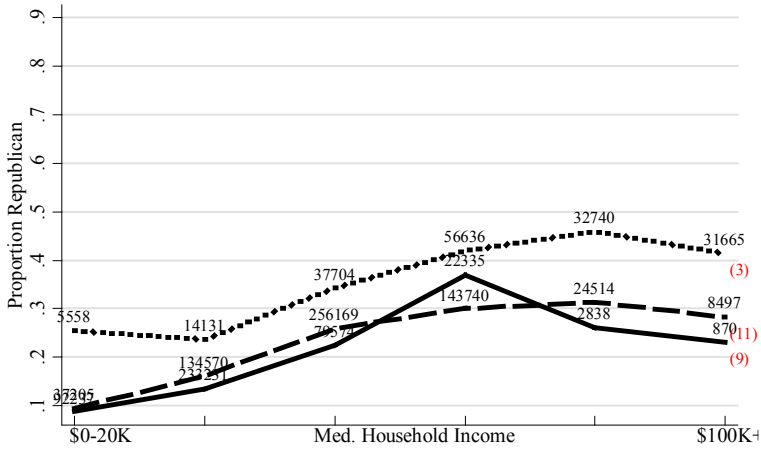
## State House Districts 0-5% Black



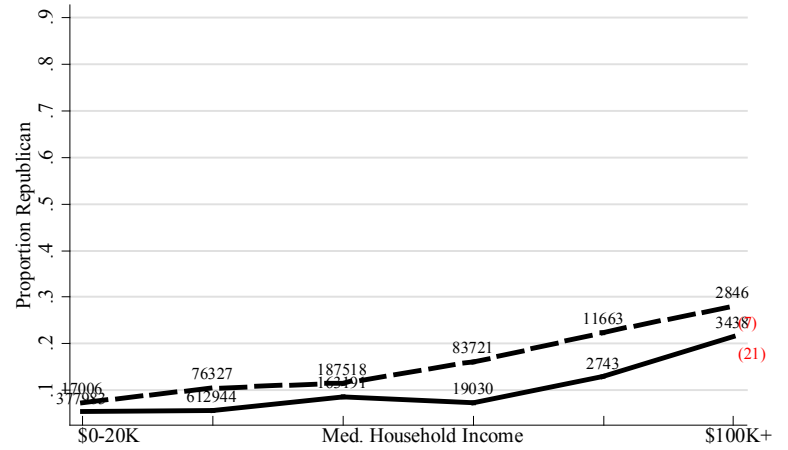
## State House Districts 5-10% Black



## State House Districts 10-25% Black

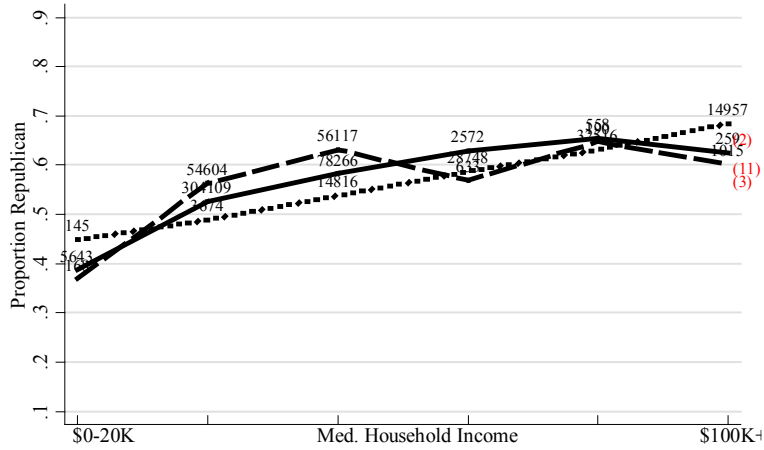


## State House Districts 25+% Black

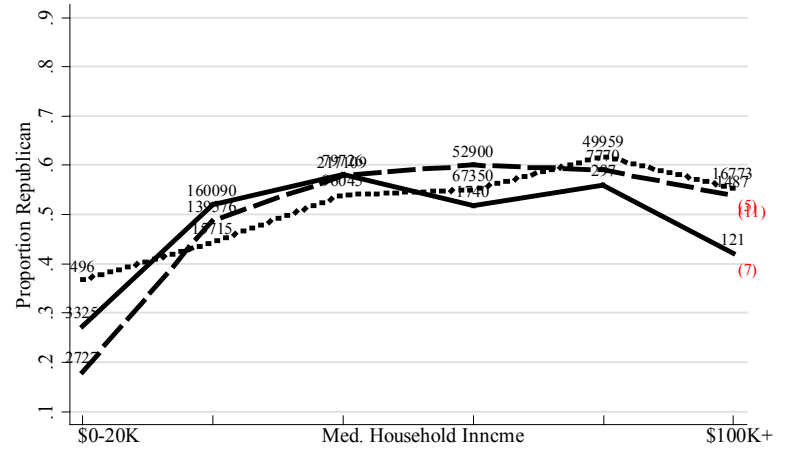


# North Carolina

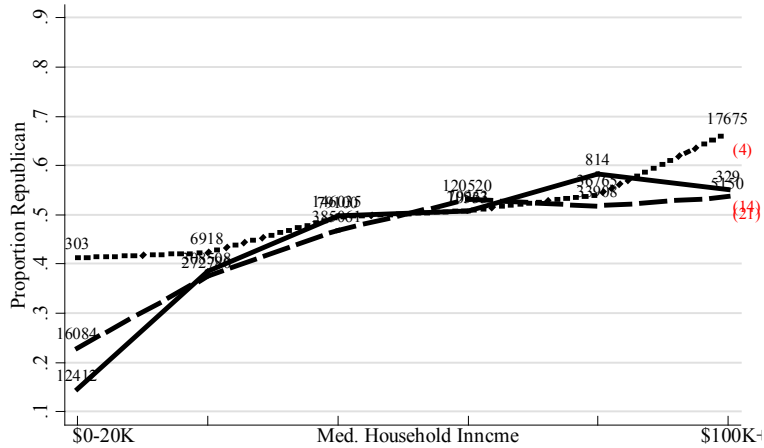
## State House Districts 0-5% Black



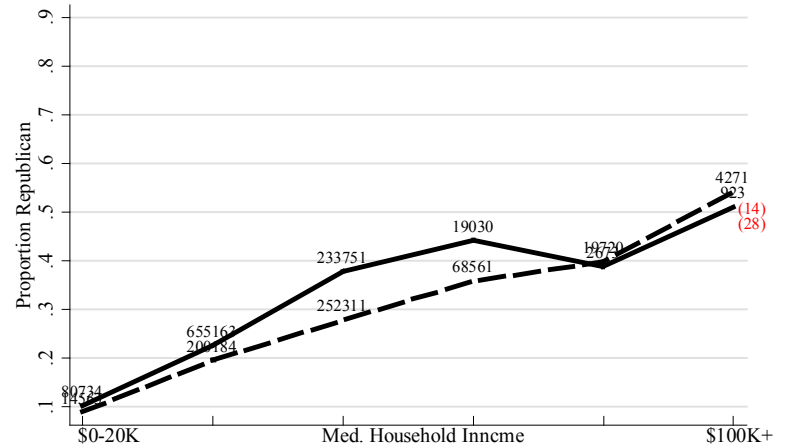
## State House Districts 5-10% Black



## State House Districts 10-25% Black



## State House Districts 25+% Black

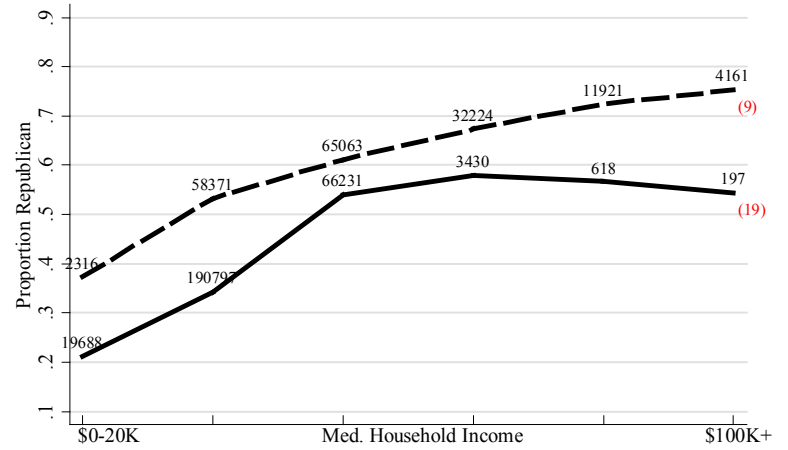


# Oklahoma

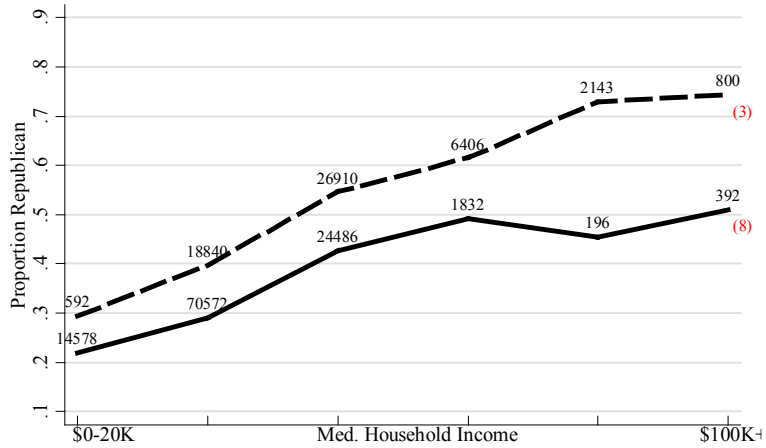
## State House Districts 0-5% Black



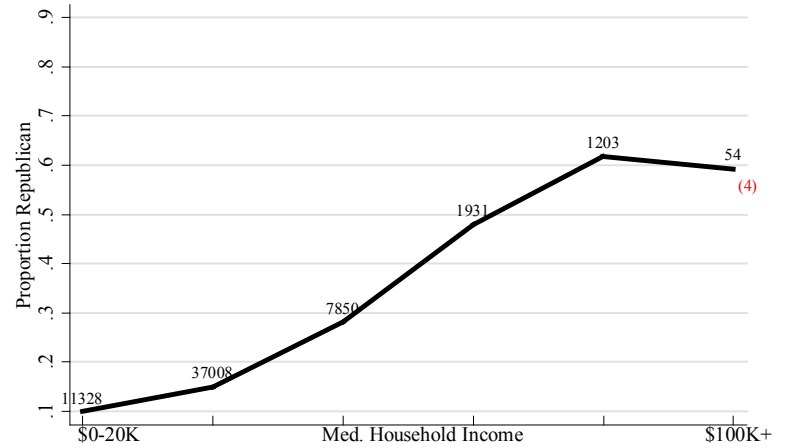
## State House Districts 5-10% Black



## State House Districts 10-25% Black

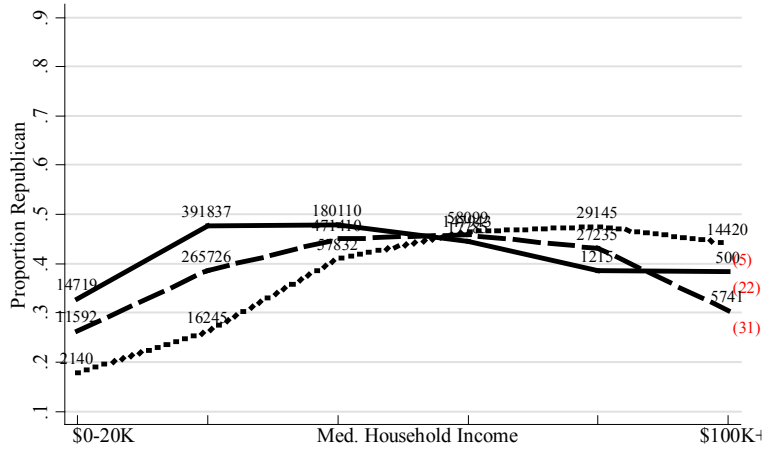


## State House Districts 25+% Black

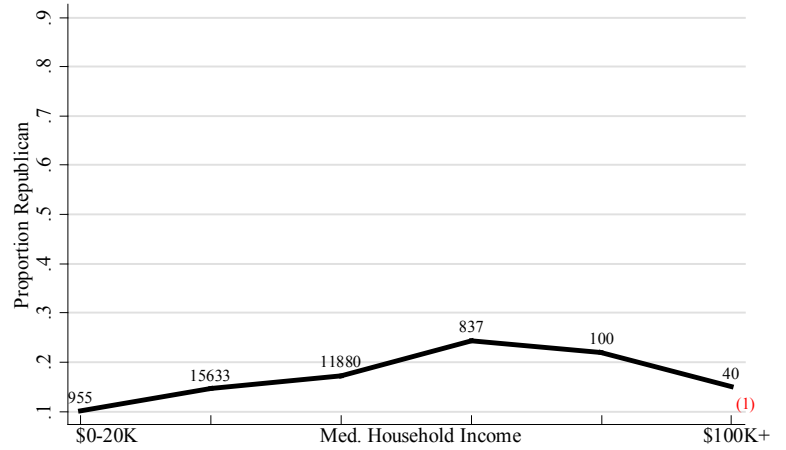


# Oregon

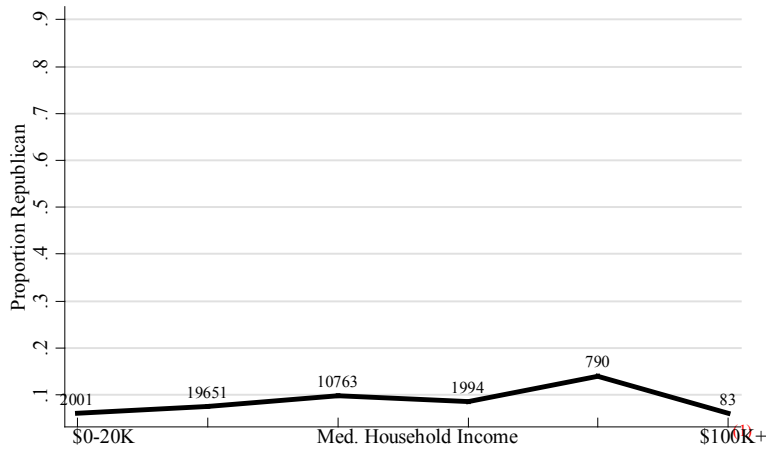
## State House Districts 0-5% Black



## State House Districts 5-10% Black

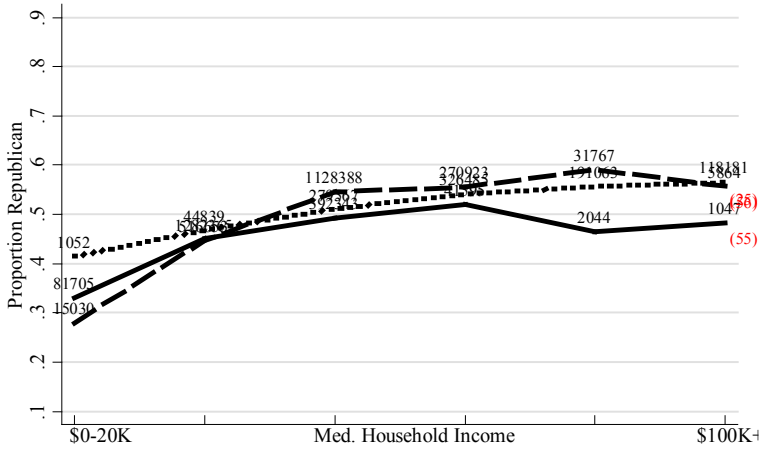


## State House Districts 25+% Black

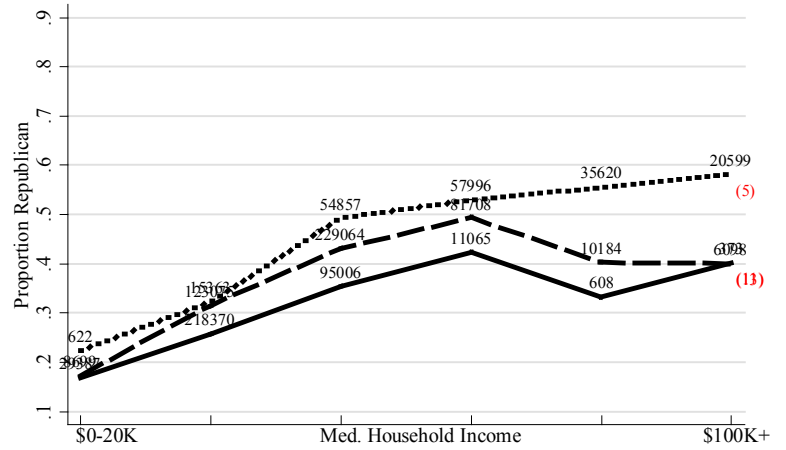


# Pennsylvania

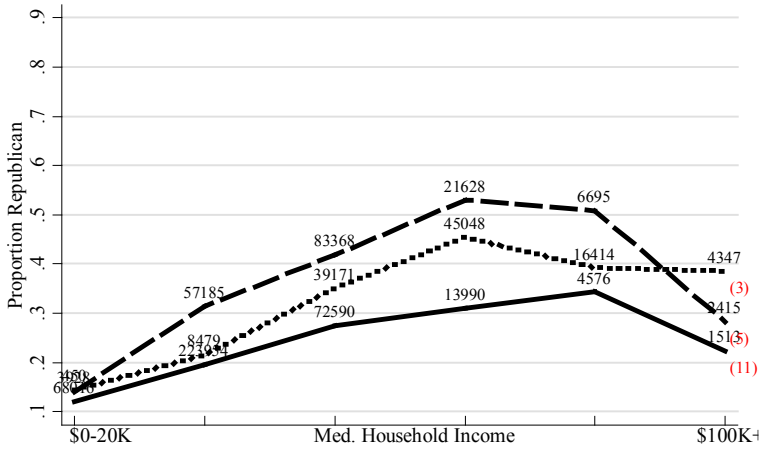
## State House Districts 0-5% Black



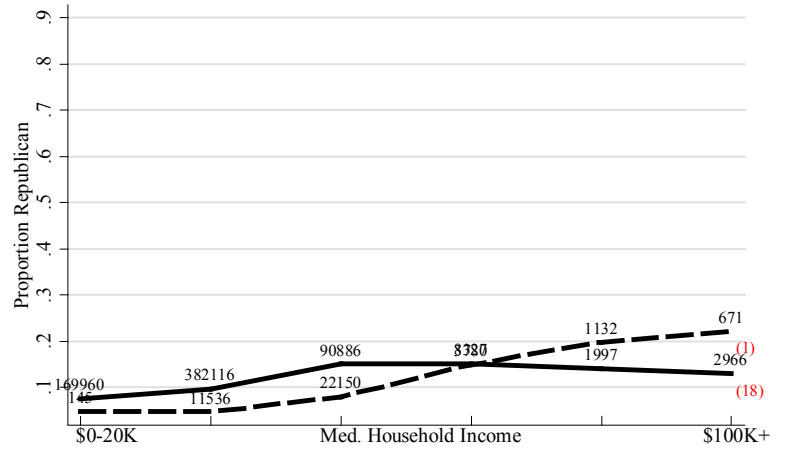
## State House Districts 5-10% Black



## State House Districts 10-25% Black



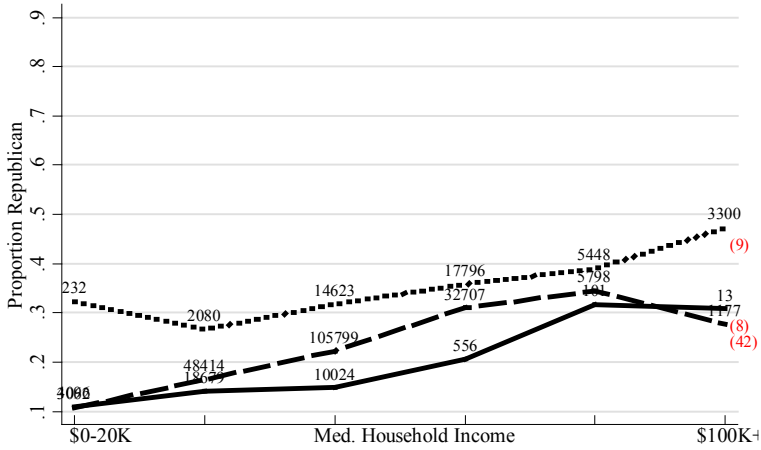
## State House Districts 25+% Black



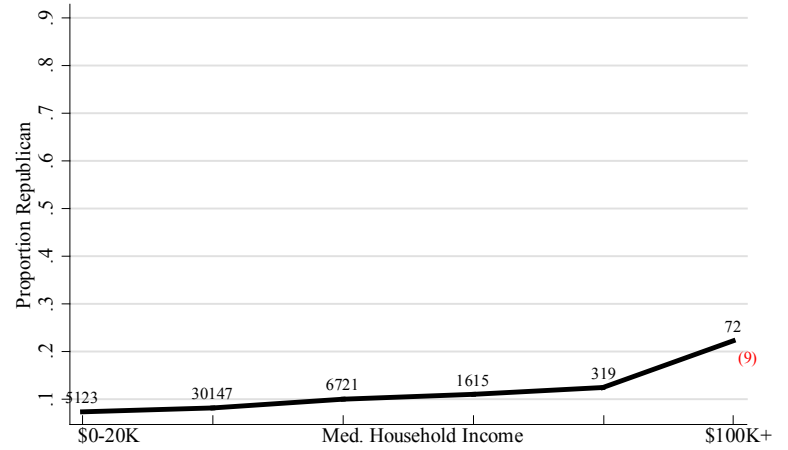


# Rhode Island

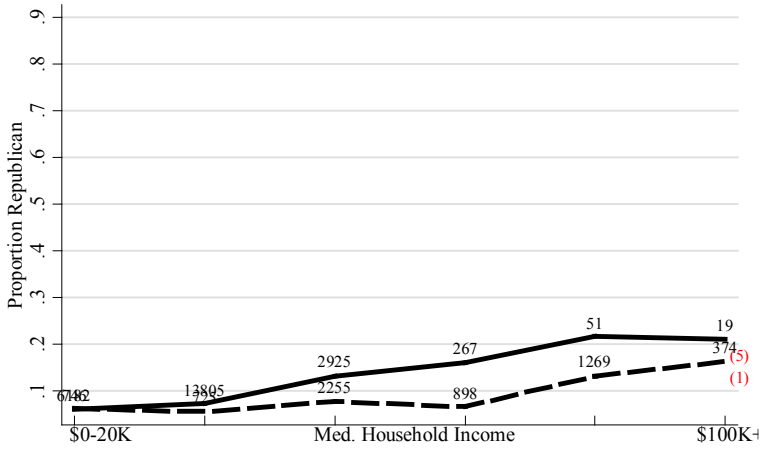
## State House Districts 0-5% Black



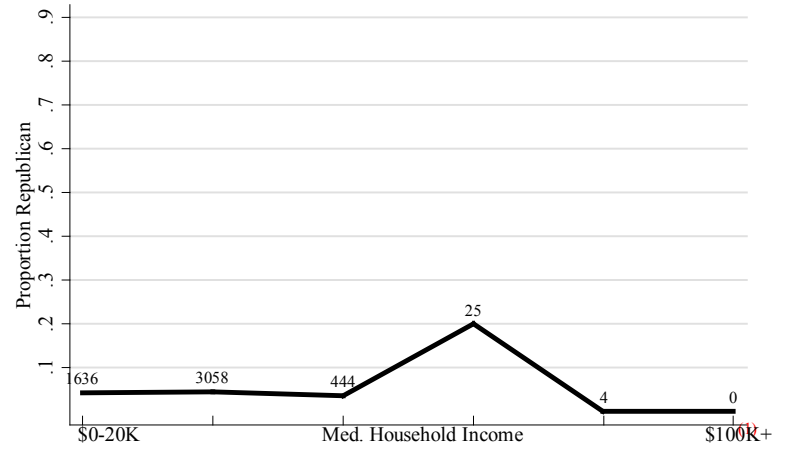
## State House Districts 5-10% Black



## State House Districts 10-25% Black

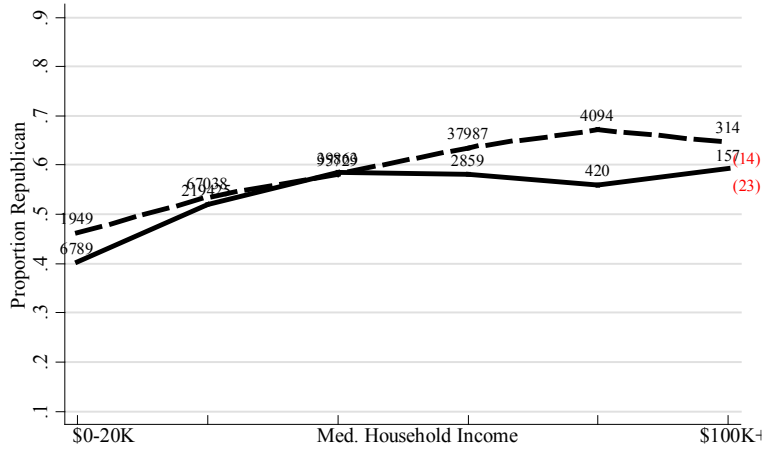


## State House Districts 25+% Black

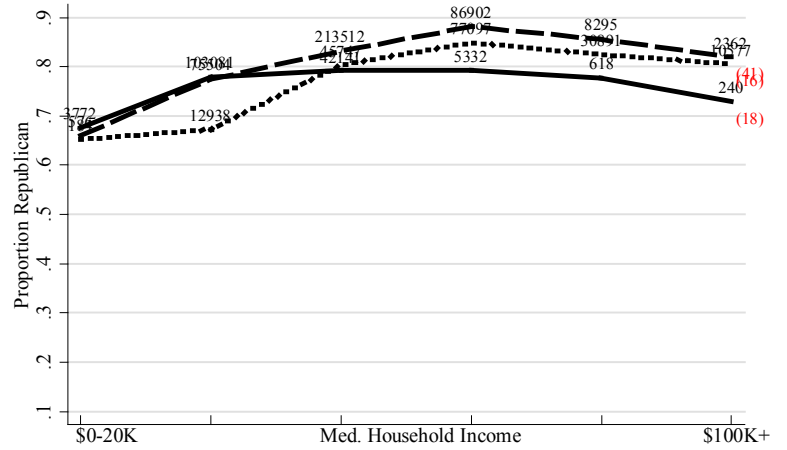


# South Dakota, Utah, and Wyoming

## SOUTH DAKOTA - State House Districts 0-5% Black



## UTAH - State House Districts 0-5% Black

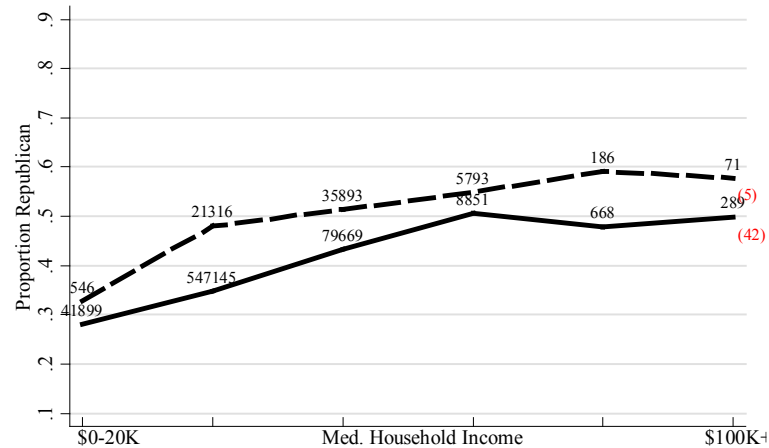


## WYOMING - State House Districts 0-5% Black

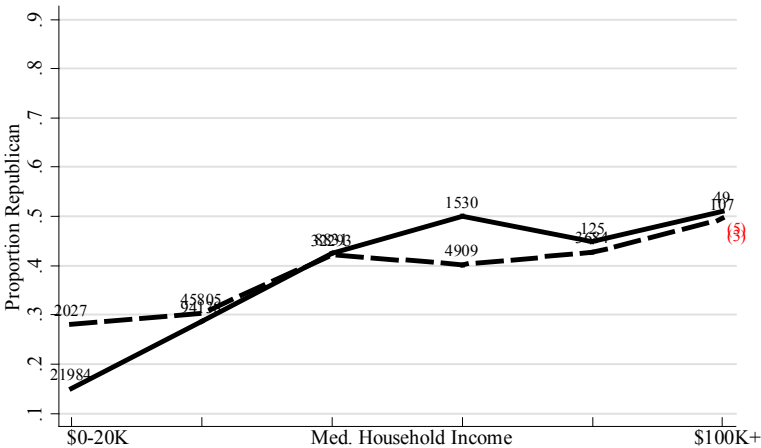


# West Virginia

State House Districts 0-5% Black



State House Districts 5-10% Black



State House Districts 10-25% Black

