

S Supplementary

S.1 Data Collection and Summary

Table S1: American Community Survey (ACS) subject tables aggregated by county. “Size” shows the 80% percentile interval of the number of housing units or population described by “Sample” (rounded to 3 significant digits). Some sizes are identical because they estimate equivalent quantities. The row marked “*” combines the data from the two indicated tables. These summary tables have been aggregated by county and racial/ethnic group, and are not the raw survey data. Further processing, such as adding noise and suppressing values to preserve anonymity of interviewees, was done on these tables by the US Census ([US Census Bureau, 2015-2020, 2021b,a](#)).

Table	Description	Sample	Size (1,000s)
S1501	Education	Population aged 18+	3.85 - 167
S2301	Employment	Population aged 16+	3.99 - 172
S2503	Financial characteristics	Occupied housing units	1.9 - 79.6
S2502	Housing demographics	Occupied housing units	1.9 - 79.6
S1903	Income	Occupied housing units	1.9 - 79.6
S2501	Housing occupancy characteristics	Occupied housing units	1.9 - 79.6
S2506+ S2507*	Housing financial characteristics	Owned housing units	1.41 - 55.2
B98001	Survey sample size	Interviewed households	0.322 - 6.44

Table S2: Summary statistics of county-level US Census American Community Survey (ACS) data. The range shows the 80% percentile interval of the non-missing values (rounded to 3 significant digits). These variables have been derived from the ACS tables using simple calculations (e.g., dividing the number of owner-occupied households by the total occupied households to get the home ownership rate).

Variable Name	Description (units)	County Range
Housing Variables		
Home ownership	Percentage of occupied households	61.5% - 81.1%
Sociodemographic Variables		
High school attainment	Percentage of 18+ population	79.6% - 93.4%
Annual income	Annual amount in current US\$	\$39,300 - \$72,300
Unemployment	Percentage of total labor force	2.4% - 8.2%
Population	Total inhabitants	4,980 - 212,000
Survey Variables		
Sample size	Total interviews	322 - 6,440

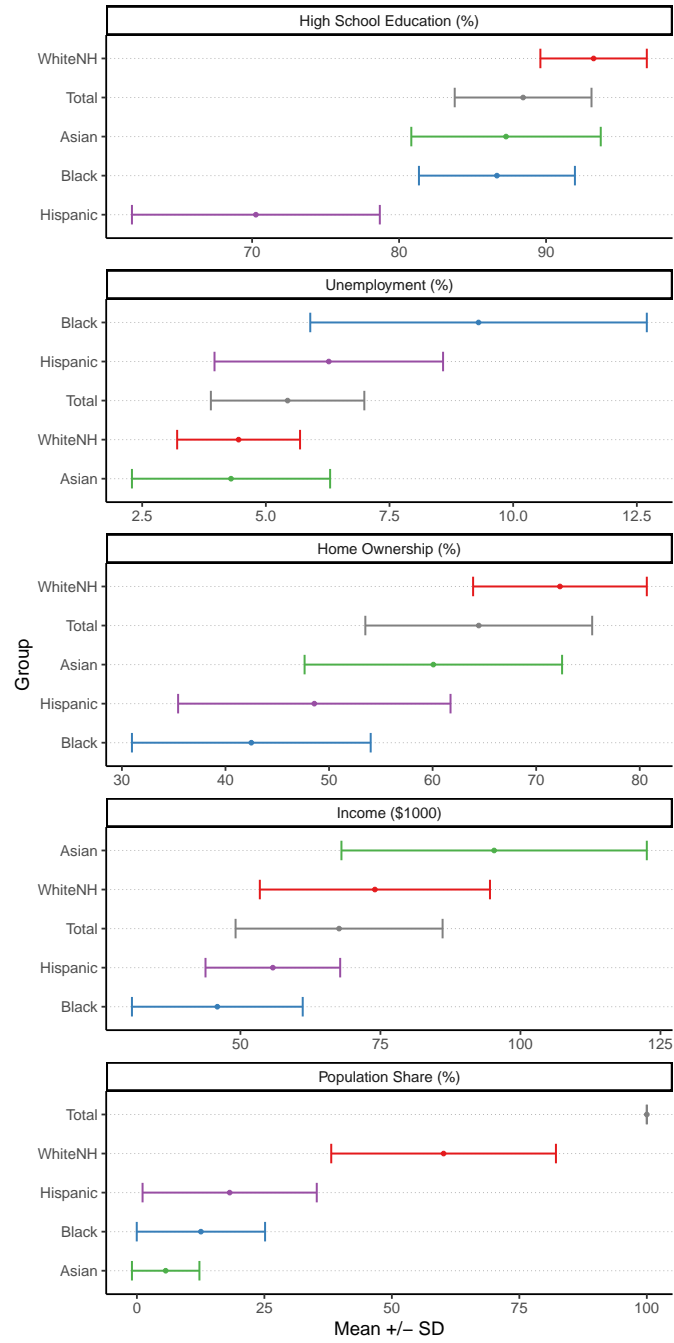


Figure S1: Dot chart showing the mean \pm SD (standard deviation) for different racial/ethnic groups and variables in the US Census American Community Survey data. The x -axes have different scales/units. The y -axes are ordered from largest mean (top) to smallest mean (bottom).

S.2 Modeling Variables and Data

Table S3: Variable names and units used for modeling. All variables other than *state* and *race* were subdivided by county and group. “Table” is the US Census table from which the quantity was derived ([US Census Bureau, 2015-2020](#)).

Name	Type	Description	Table
<i>state</i>	Categorical	50 US states.	N/A
<i>group</i>	Categorical	WhiteNH (White non-Hispanic), Hispanic, Black, or Asian.	N/A
<i>ownersp</i>	Numerical	Proportion of home owners.	S2502
<i>owners</i>	Integer	Number of owned households.	S2502, B98001
<i>sampsize</i>	Integer	Number of interviewed households.	DP05, B98001
<i>hsedu</i>	Numerical	High-school completion proportion.	S1501
<i>income</i>	Numerical	Household income in past 12 months in units of \$100,000.	S1903
<i>unemp</i>	Numerical	Unemployment proportion.	S2301
<i>logpop</i>	Numerical	Log total number of households.	DP05

Table S4: The percentages of missing values in each variable (rounded to two decimals). The two values separated by a “/” are the percentages weighted with and without, respectively, the sample size of the corresponding observation. The percentages are calculated separately for each group and variable as indicated by the row names and column names, respectively. The “All” column shows observations that have any of the variables missing, while the “All” row shows the sum of the missing observations over all the groups.

	<i>ownersp</i>	<i>income</i>	<i>hsedu</i>	<i>unemp</i>	All
WhiteNH	0.00/00.00	0.01/00.22	0.00/00.00	0.00/00.00	0.01/00.22
Hispanic	0.01/02.26	0.80/21.93	0.00/00.92	0.01/01.69	0.80/21.99
Black	0.16/14.83	1.31/40.17	0.00/04.33	0.08/09.87	1.31/40.23
Asian	0.12/18.84	1.99/56.84	0.01/09.61	0.05/14.00	1.99/56.84
All	0.03/08.98	0.38/29.79	0.00/03.72	0.01/06.39	0.38/29.82

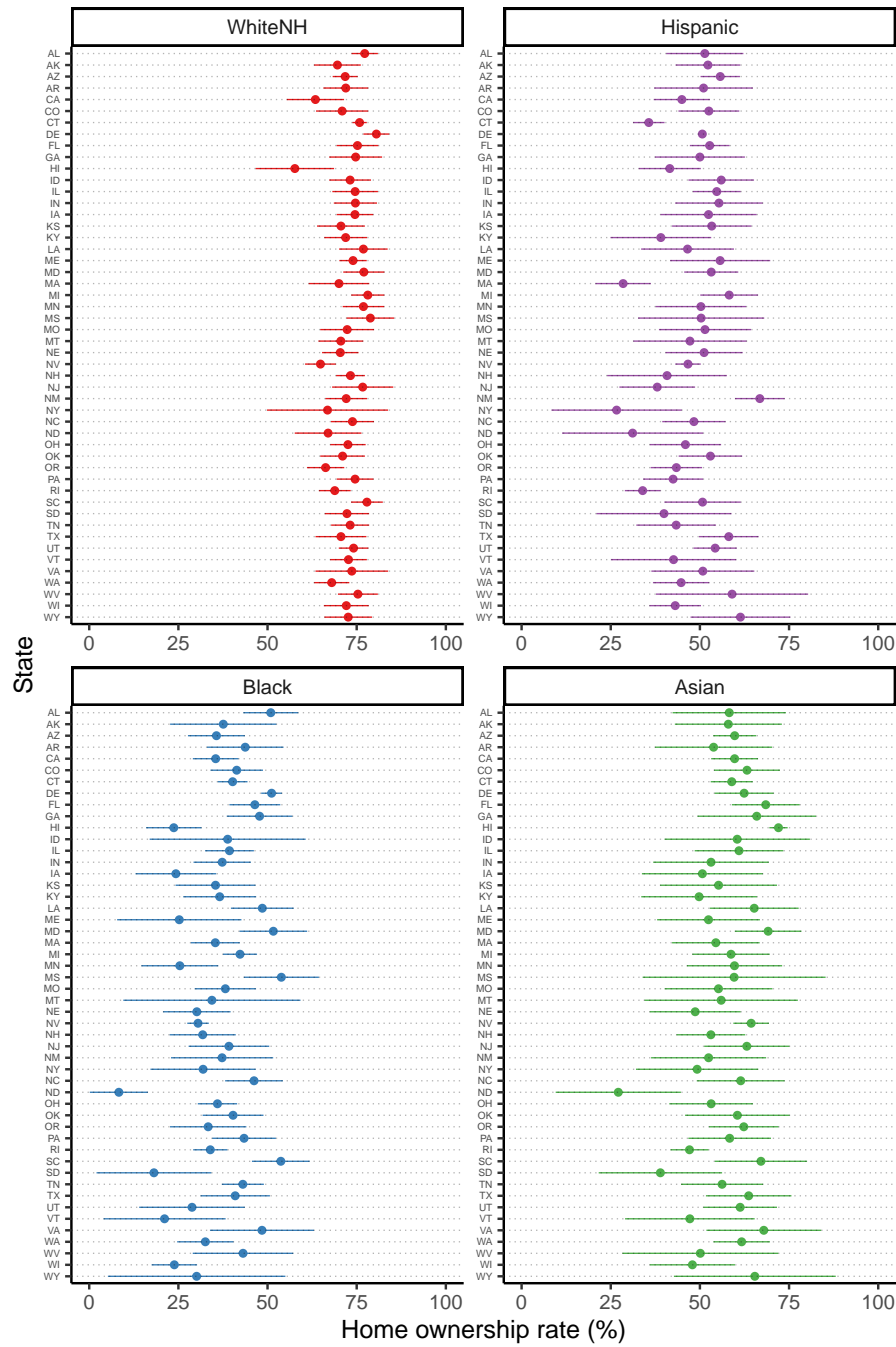
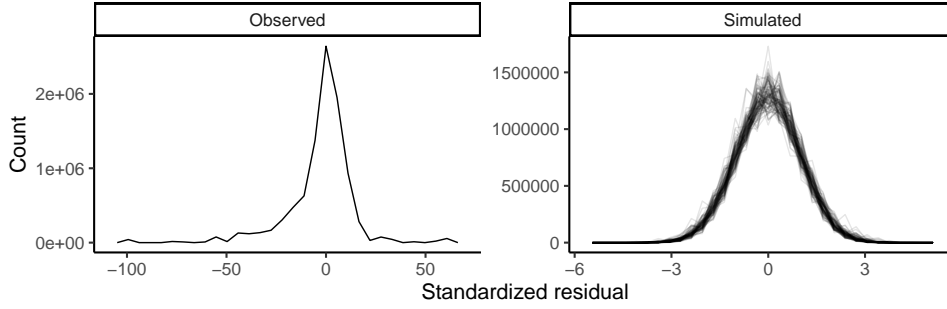
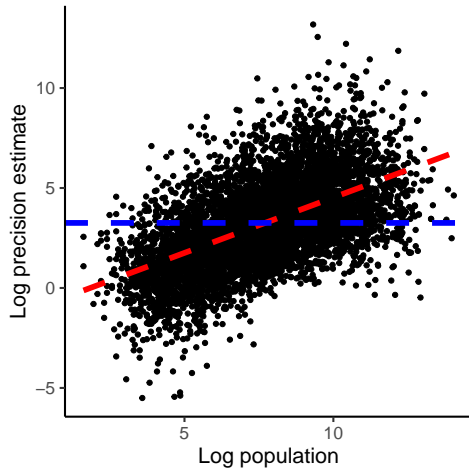


Figure S2: Point-interval plot summarizing the home ownership rate in each state. The mean is shown as a point and the mean \pm SD (standard deviation) is shown as a line segment. The mean and SD are weighted by the populations of the counties within the state. The y -axis shows the state (using the official abbreviations) and the x -axis shows the home ownership rate. There is a significant difference in the home ownership rate of different states. Made with the ggplot2 (Wickham, 2016) and ggdist (Kay, 2023) R packages.

S.3 Binomial Model Analysis



(a) Comparison of the Pearson residuals of the observed data (left) with the residuals of the replicate data sets (right) using the fitted binomial model. The kernel density estimate weighted by the sample size of each observation is shown. The left and right panels have different x - and y -axis scales. We see high over-dispersion in the observed residuals.



(b) The points $\{(\logpop_k, z_k)\}_k$ and the fitted least squares line $z = -1.02 + .55\logpop$ (red) and mean line $z = 3.25$ (blue). Each point represents a single observation with a properly defined value for z_k . As expected, there is a positive association between \logpop and z , since z estimates $\log(\phi)$.

Figure S3: Binomial model analysis.

In our exploratory modeling we fit a preliminary model with a binomial response using maximum likelihood estimation (MLE) with the base R function `glm` (R Core Team, 2021). The formulation of this model is given in Section 3.2 (main text) as Model 0. We observed that the binomial model was a poor fit for our data. To show this, we used the fitted coefficients of the binomial model to simulate replicate data sets and compare the distribution of the Pearson residuals to the same residuals of the data. We define the Pearson residuals as in (Agresti, 2012, ch 6.2.1) by $\varepsilon_k = (y_k - N_k \hat{p}_k) / \sqrt{N_k(1 - \hat{p}_k)\hat{p}_k}$, where y_k is the observed number of owned households, N_k is the sample size, and \hat{p}_k is the fitted probability of home ownership in the binomial model. Figure S3a shows the density comparison of the simulated residuals vs. the observed residuals. We see that the observed residuals are highly over-dispersed, taking on values at low as -100

and as high as 50, while the simulated residuals are typically within the range $[-5, 5]$. This shortcoming of the binomial model was our initial motivation for using the beta-binomial, with the additional flexibility allowed by its precision component.

We also used the fitted probabilities of the binomial model to estimate plausible values for the precision component. To do so, we followed a procedure similar to diagnosing non-constant variance in ordinary least squares model by plotting the residuals as a function of the fitted values as in (Draper and Smith, 1998, p. 62). First, let y_k and N_k be the observed number of owned households and sample size, respectively, where k indexes each observation. Let \hat{p}_k be the fitted probabilities from the binomial model. Assuming that \hat{p}_k would be approximately the same for a beta-binomial model and ϕ_k were the true precision parameters, we would have

$$\begin{aligned} E((y_k - \hat{p}_k N_k)^2 | \hat{p}_k) &= \text{Var}(y_k | \hat{p}_k) \approx N_k(1 - \hat{p}_k)\hat{p}_k \frac{\phi_k + N_k}{\phi_k + 1} \implies \\ E\left(\frac{(y_k - \hat{p}_k N_k)^2}{N_k(1 - \hat{p}_k)\hat{p}_k} \middle| \hat{p}_k\right) &\approx \frac{\phi_k + N_k}{\phi_k + 1} \end{aligned}$$

For brevity, define

$$r_k = \frac{(y_k - \hat{p}_k N_k)^2}{N_k(1 - \hat{p}_k)\hat{p}_k}$$

Then we get

$$\begin{aligned} E(r_k | \hat{p}_k) &\approx \frac{\phi_k + N_k}{\phi_k + 1} \implies \\ \frac{N_k - 1}{E(r_k | \hat{p}_k) - 1} &\approx \phi_k \end{aligned}$$

We make an approximation by bringing the expectation out to get

$$E\left(\frac{N_k - 1}{r_k - 1} \middle| \hat{p}_k\right) \approx \phi_k \quad (\text{S1})$$

In all our beta-binomial models, $\log(\phi_k)$ is modeled as intercept-only or as a linear function of the log population, $\log pop_k$. We use the latter formulation here to estimate the linear-scale coefficients:

$$E(\log(\phi_k)) = \beta_0 + \beta_1 \log pop_k \quad (\text{S2})$$

Using (S1) and (S2) we get

$$\log\left(E\left(\frac{N_k - 1}{r_k - 1} \middle| \hat{p}_k\right)\right) \approx \beta_0 + \beta_1 \log pop_k$$

We make another approximation by exchanging the log and the expectation to get

$$E\left(\log\left(\frac{N_k - 1}{r_k - 1}\right) \middle| \hat{p}_k\right) \approx \beta_0 + \beta_1 \log pop_k \quad (\text{S3})$$

Then, to estimate β_0 and β_1 , we fit an ordinary least squares line to the points

$$\{(\log pop_k, z_k)\}_k$$

where

$$z_k = \log\left(\frac{N_k - 1}{r_k - 1}\right) \approx \log(\phi_k)$$

Some z_k values were undefined, such as when $N_k = 1$ or $r_k \leq 1$, in which case the point was omitted. Figure S3b shows the points and the fitted line.

The estimated coefficients from the line were $\hat{\beta}_0 = -1.02$ and $\hat{\beta}_1 = 0.55$. The estimate for $\hat{\beta}_1$ was in line with our expectation that precision should increase with population. The mean of the z_k was about 3.25, which is our prior mean for the Intercept parameter of precision.

S.4 Model Selection

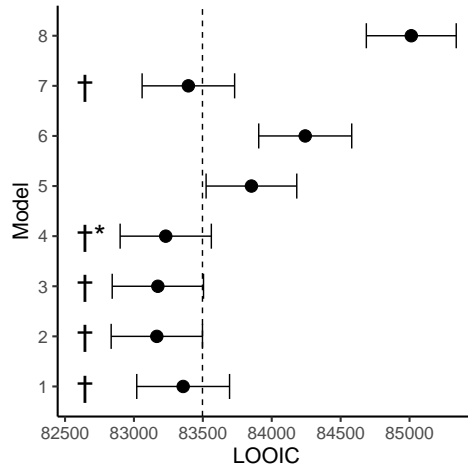
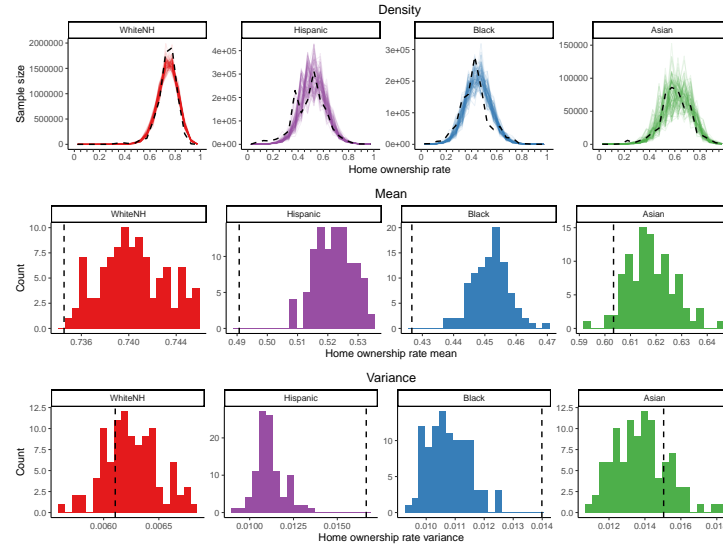
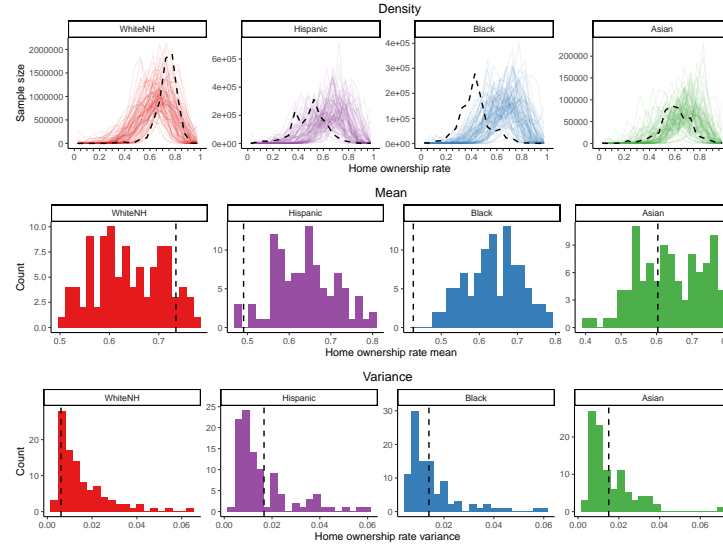


Figure S4: The LOOIC and corresponding standard error (SE) for each model. The asterisk (*) denotes the final selected model while the dagger (†) denotes all models with LOOIC within 1 SE of the minimum (Model 2). Lower LOOIC indicates better predictive performance. Although Model 7 was the most parsimonious model within 1 SE of the minimum, we chose Model 4 due to finding evidence that interaction effects were important for the model (see Section 3.4 (main text)).

S.5 Model Diagnostics



(a) Posterior predictive plots.



(b) Prior predictive plots.

Figure S5: Posterior (a) and prior (b) predictive plots for Model 4. In the “Density” plot, the colored lines represent histograms of home ownership rate for the posterior (or prior) replicates, and the black dashed line is the histogram of home ownership rate for the data. In the “Mean” and “Variance” plots, the colored bars show the histogram of the corresponding summary statistic for the replicates, and the black dashed vertical line shows the summary statistic for the data. The x -axis in all panels is the home ownership rate (different scales used) and the y -axis is the total sample size in each bin (different scales used). The histograms and summary statistics are weighted by the sample size of each observation. 100 draws are used.

Figure S5a shows the posterior predictive checks for Model 4. Although the response distribution for the beta-binomial is a count, we converted all counts to proportions by dividing

by the sample sizes. Thus, the predictive checks are made on the home ownership rate, rather than the number of owned households. The plot in Figure S5a titled “Density” shows the empirical densities of the posterior draws compared to the data. While the WhiteNH and Asian data appear to be consistent with the posterior densities, the Hispanic and Black groups contain significant deviations from the posterior replicates. In particular, we see a secondary mode in the Hispanic group home ownership density around 0.3-0.4, which is not encompassed by the posterior replicate densities. This mode appears to be a result of California counties, including Los Angeles County, Orange County, Santa Clara County, and Alameda County, which have a high population of Hispanic inhabitants and a relatively low home ownership rate, possibly due to being highly urban. In the Black group, we also see a secondary mode around 0.55-0.7, which may be a result of Prince Georges County, Maryland, which has a relatively large Black population (62% of the county) and a relatively high home ownership rate (61%) for the Black group. There are other deviations in the data, which indicate that the model is not a perfect fit.

In the “Mean” and “Variance” plots of Figure S5a, we see the corresponding weighted summary statistics of the data compared to the histogram of the posterior replicate summary statistics. The WhiteNH and Asian posterior distributions appear to fit the data marginally well, while the Hispanic and Black appear to fit poorly. Further modifications to the model may be required to resolve these discrepancies.

We also show the prior predictive check in Figure S5b. See Section 3 (main text) for an explanation of how this prior was selected. While there are clear deviations between the prior and empirical densities, we see that they generally overlap well. This is also true for the prior distribution of the mean and variance. We consider this acceptable as our prior is intended to be only weakly informative.

S.6 Model Coefficients

Table S5: Posterior coefficient estimates for Model 4 (including the state indicator variables). “Mean” is the posterior mean, “SD” is the standard deviation, and “2.5%” and “97.5%” are the corresponding quantiles. The notation “A:B” is the interaction term between the variables “A” and “B”. The parameter θ is the home ownership rate and ϕ is the precision.

(a) Coefficients at log-odds (for θ) and log (for ϕ) scale. (b) Coefficients at odds-ratio (for θ) and ratio (for ϕ) scale.

Component	Name	Mean	SD	2.5%	97.5%
logit(θ)	Intercept	1.03	0.04	0.96	1.11
logit(θ)	Hispanic	-1.28	0.05	-1.37	-1.18
logit(θ)	Black	-1.47	0.04	-1.56	-1.39
logit(θ)	Asian	-1.20	0.06	-1.31	-1.08
logit(θ)	income:WhiteNH	0.26	0.05	0.16	0.35
logit(θ)	income:Hispanic	0.91	0.07	0.78	1.05
logit(θ)	income:Black	0.76	0.07	0.62	0.89
logit(θ)	income:Asian	0.97	0.06	0.86	1.09
logit(θ)	Alaska	-0.51	0.08	-0.66	-0.36
logit(θ)	Arizona	-0.06	0.06	-0.18	0.06
logit(θ)	Arkansas	-0.04	0.05	-0.13	0.06
logit(θ)	California	-0.45	0.04	-0.52	-0.38
logit(θ)	Colorado	-0.18	0.05	-0.27	-0.08
logit(θ)	Connecticut	-0.36	0.07	-0.49	-0.23
logit(θ)	Delaware	0.06	0.11	-0.14	0.27
logit(θ)	Florida	0.09	0.04	0.01	0.16
logit(θ)	Georgia	0.04	0.04	-0.03	0.10
logit(θ)	Hawaii	-0.33	0.10	-0.52	-0.13
logit(θ)	Idaho	-0.09	0.06	-0.21	0.03
logit(θ)	Illinois	-0.12	0.04	-0.20	-0.04
logit(θ)	Indiana	-0.12	0.04	-0.20	-0.04
logit(θ)	Iowa	-0.18	0.04	-0.27	-0.10
logit(θ)	Kansas	-0.19	0.05	-0.28	-0.10
logit(θ)	Kentucky	-0.22	0.04	-0.30	-0.14
logit(θ)	Louisiana	0.13	0.04	0.05	0.22
logit(θ)	Maine	-0.13	0.08	-0.28	0.02
logit(θ)	Maryland	-0.10	0.05	-0.20	0.01
logit(θ)	Massachusetts	-0.51	0.06	-0.62	-0.40
logit(θ)	Michigan	0.07	0.04	-0.01	0.15
logit(θ)	Minnesota	-0.10	0.04	-0.18	-0.01
logit(θ)	Mississippi	0.29	0.05	0.20	0.37
logit(θ)	Missouri	-0.22	0.04	-0.30	-0.14
logit(θ)	Montana	-0.24	0.06	-0.36	-0.11
logit(θ)	Nebraska	-0.23	0.05	-0.33	-0.13
logit(θ)	Nevada	-0.27	0.07	-0.42	-0.13
logit(θ)	NewHampshire	-0.32	0.08	-0.48	-0.16
logit(θ)	NewJersey	-0.29	0.05	-0.38	-0.19
logit(θ)	NewMexico	0.16	0.06	0.04	0.28
logit(θ)	NewYork	-0.46	0.04	-0.54	-0.39
logit(θ)	NorthCarolina	-0.01	0.04	-0.08	0.06
logit(θ)	NorthDakota	-0.44	0.06	-0.57	-0.32
logit(θ)	Ohio	-0.20	0.04	-0.27	-0.13
logit(θ)	Oklahoma	-0.02	0.05	-0.11	0.06
logit(θ)	Oregon	-0.36	0.05	-0.47	-0.26
logit(θ)	Pennsylvania	-0.21	0.04	-0.29	-0.13
logit(θ)	RhodeIsland	-0.47	0.10	-0.66	-0.29
logit(θ)	SouthCarolina	0.22	0.05	0.13	0.31
logit(θ)	SouthDakota	-0.24	0.06	-0.36	-0.12
logit(θ)	Tennessee	-0.13	0.04	-0.21	-0.05
logit(θ)	Texas	0.09	0.03	0.03	0.16
logit(θ)	Utah	-0.12	0.06	-0.24	0.01
logit(θ)	Vermont	-0.17	0.09	-0.34	0.01
logit(θ)	Virginia	-0.09	0.04	-0.16	-0.01
logit(θ)	Washington	-0.31	0.05	-0.40	-0.21
logit(θ)	WestVirginia	0.02	0.06	-0.10	0.13
logit(θ)	Wisconsin	-0.30	0.04	-0.38	-0.22
logit(θ)	Wyoming	-0.08	0.08	-0.24	0.08
log(ϕ)	Intercept	0.10	0.06	-0.02	0.21
log(ϕ)	logpop	0.34	0.01	0.32	0.35

Component	Name	Mean	SD	2.5%	97.5%
$\theta/(1-\theta)$	Intercept	2.81	0.11	2.60	3.04
$\theta/(1-\theta)$	Hispanic	0.28	0.01	0.25	0.31
$\theta/(1-\theta)$	Black	0.23	0.01	0.21	0.25
$\theta/(1-\theta)$	Asian	0.30	0.02	0.27	0.34
$\theta/(1-\theta)$	income:WhiteNH	1.29	0.06	1.18	1.42
$\theta/(1-\theta)$	income:Hispanic	2.50	0.17	2.18	2.86
$\theta/(1-\theta)$	income:Black	2.14	0.15	1.87	2.44
$\theta/(1-\theta)$	income:Asian	2.65	0.16	2.35	2.97
$\theta/(1-\theta)$	Alaska	0.60	0.05	0.51	0.69
$\theta/(1-\theta)$	Arizona	0.94	0.06	0.84	1.06
$\theta/(1-\theta)$	Arkansas	0.96	0.04	0.88	1.06
$\theta/(1-\theta)$	California	0.64	0.02	0.60	0.69
$\theta/(1-\theta)$	Colorado	0.84	0.04	0.77	0.92
$\theta/(1-\theta)$	Connecticut	0.70	0.05	0.61	0.80
$\theta/(1-\theta)$	Delaware	1.07	0.11	0.87	1.31
$\theta/(1-\theta)$	Florida	1.09	0.04	1.01	1.17
$\theta/(1-\theta)$	Georgia	1.04	0.04	0.97	1.11
$\theta/(1-\theta)$	Hawaii	0.72	0.07	0.59	0.88
$\theta/(1-\theta)$	Idaho	0.92	0.06	0.81	1.03
$\theta/(1-\theta)$	Illinois	0.89	0.04	0.82	0.96
$\theta/(1-\theta)$	Indiana	0.89	0.04	0.82	0.96
$\theta/(1-\theta)$	Iowa	0.83	0.04	0.76	0.91
$\theta/(1-\theta)$	Kansas	0.83	0.04	0.75	0.90
$\theta/(1-\theta)$	Kentucky	0.80	0.03	0.74	0.87
$\theta/(1-\theta)$	Louisiana	1.14	0.05	1.05	1.24
$\theta/(1-\theta)$	Maine	0.88	0.07	0.76	1.02
$\theta/(1-\theta)$	Maryland	0.91	0.05	0.82	1.01
$\theta/(1-\theta)$	Massachusetts	0.60	0.03	0.54	0.67
$\theta/(1-\theta)$	Michigan	1.07	0.04	0.99	1.16
$\theta/(1-\theta)$	Minnesota	0.91	0.04	0.83	0.99
$\theta/(1-\theta)$	Mississippi	1.33	0.06	1.22	1.45
$\theta/(1-\theta)$	Missouri	0.80	0.03	0.74	0.87
$\theta/(1-\theta)$	Montana	0.79	0.05	0.70	0.90
$\theta/(1-\theta)$	Nebraska	0.80	0.04	0.72	0.88
$\theta/(1-\theta)$	Nevada	0.76	0.06	0.66	0.88
$\theta/(1-\theta)$	NewHampshire	0.73	0.06	0.62	0.85
$\theta/(1-\theta)$	NewJersey	0.75	0.04	0.69	0.82
$\theta/(1-\theta)$	NewMexico	1.17	0.07	1.04	1.32
$\theta/(1-\theta)$	NewYork	0.63	0.02	0.59	0.68
$\theta/(1-\theta)$	NorthCarolina	0.99	0.04	0.92	1.06
$\theta/(1-\theta)$	NorthDakota	0.64	0.04	0.57	0.73
$\theta/(1-\theta)$	Ohio	0.82	0.03	0.76	0.88
$\theta/(1-\theta)$	Oklahoma	0.98	0.04	0.89	1.07
$\theta/(1-\theta)$	Oregon	0.70	0.04	0.63	0.77
$\theta/(1-\theta)$	Pennsylvania	0.81	0.03	0.75	0.88
$\theta/(1-\theta)$	RhodeIsland	0.63	0.06	0.52	0.75
$\theta/(1-\theta)$	SouthCarolina	1.24	0.06	1.13	1.36
$\theta/(1-\theta)$	SouthDakota	0.79	0.05	0.70	0.89
$\theta/(1-\theta)$	Tennessee	0.88	0.04	0.81	0.95
$\theta/(1-\theta)$	Texas	1.10	0.04	1.03	1.17
$\theta/(1-\theta)$	Utah	0.89	0.06	0.79	1.01
$\theta/(1-\theta)$	Vermont	0.85	0.08	0.71	1.01
$\theta/(1-\theta)$	Virginia	0.92	0.03	0.85	0.99
$\theta/(1-\theta)$	Washington	0.73	0.04	0.67	0.81
$\theta/(1-\theta)$	WestVirginia	1.02	0.06	0.91	1.14
$\theta/(1-\theta)$	Wisconsin	0.74	0.03	0.68	0.81
$\theta/(1-\theta)$	Wyoming	0.92	0.08	0.78	1.08
ϕ	Intercept	1.10	0.07	0.98	1.24
ϕ	logpop	1.40	0.01	1.38	1.42

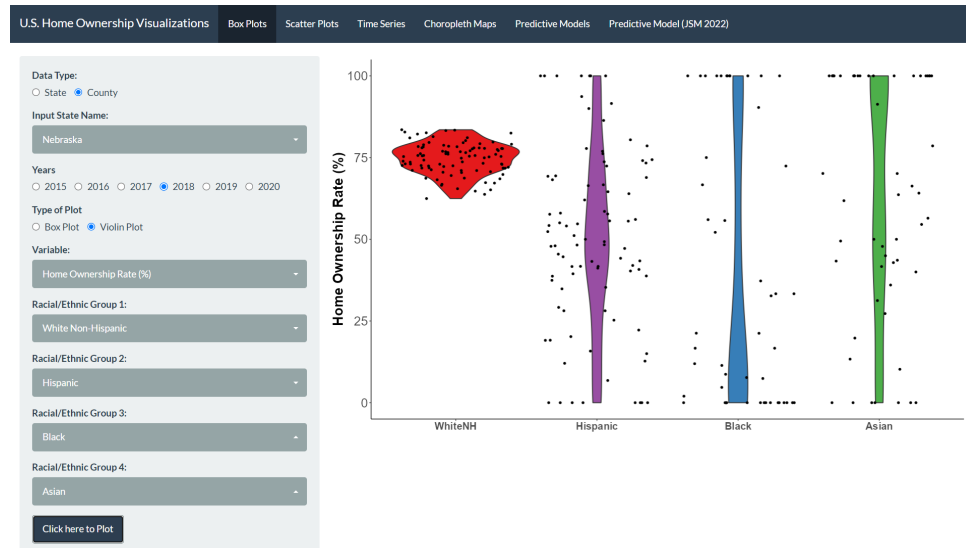
S.7 Shiny App

We developed a web interface using the R package Shiny (Chang et al., 2022) and shinyWidgets (Perrier et al., 2023), so that users can run customized and dynamic explorations of the data and models presented here.

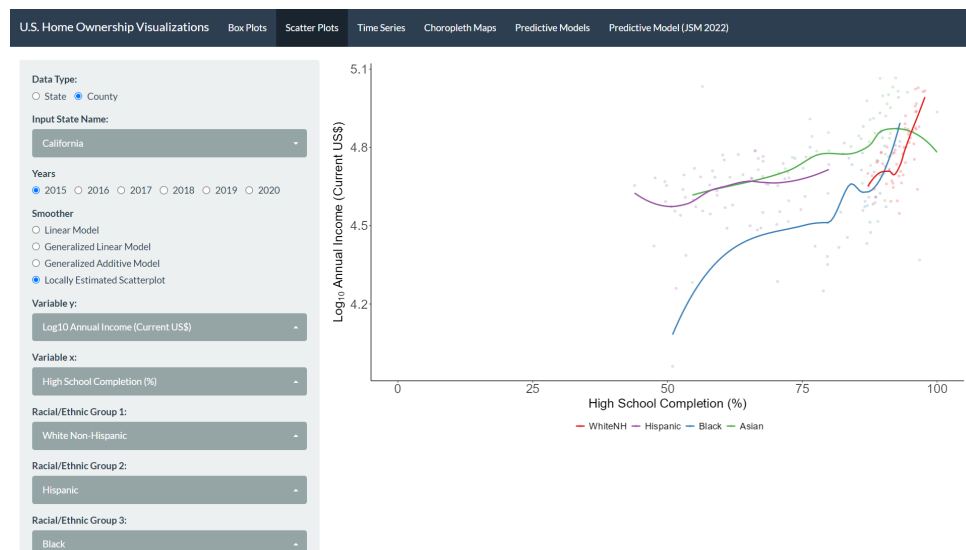
This Shiny app code can be accessed at

https://github.com/jmedri/JSM2022_HomeOwnership. This app currently has five menu options.

- (a) **Box Plots:** Users can generate a customized box plot, as depicted in Figure 2a (main text). Users have the flexibility to choose the data type (aggregated by state or county), the desired year (ranging from 2015 to 2020), the plot type (either box plot or violin plot), the variable of interest (housing and sociodemographic variables), and up to four racial/ethnic groups for comparison.
- (b) **Scatter Plots:** Users can create a smoothed scatter plot for two variables, like the ones shown in Figures 1a and 1b (main text). Users have the flexibility to choose the data type (aggregated by state or county), the desired year (ranging from 2015 to 2020), the smoother type (linear model, generalized linear model, generalized additive model, and locally estimated scatterplot), the x - and y -variable of interest (housing and sociodemographic variables), and up to four racial/ethnic groups for comparison.
- (c) **Time Series Plots:** Users can produce a time series plot. Users have the flexibility to choose the data type (aggregated by state or county), the variable of interest (housing and sociodemographic variables), and up to four racial/ethnic groups for comparison.
- (d) **Choropleth Maps:** Users can generate choropleth maps, similar to the example displayed in Figure 2b (main text). This feature provides the flexibility to select the data type, allowing visualization of either the entire nation or a specific state. Additionally, users can choose the desired year (ranging from 2015 to 2020), select the variable of interest (housing and sociodemographic variables), customize the color palette, and compare up to four racial/ethnic groups.
- (e) **Predictive Models:** With this feature, users can use any of the nine models discussed in Section 3 (main text) to predict home ownership rates using aggregated county data. Furthermore, users have the option to select up to four racial/ethnic groups, specify a particular state (or view nationwide predictions), and input specific values for each variable based on their chosen model. The right-hand plot provides a visual representation of the predicted values for the selected model, including density, median, and a 95% predictive interval, which can be displayed either separately or overlaid together.
- (f) **Predictive Models (JSM 2022):** This feature shows the predicted home ownership rates of the models showcased in the Data Expo Challenge at JSM 2022 (Medri and Channagiri, 2022), which uses aggregated state data. Users can select from a range of model types, including Maximum Likelihood Estimation Gaussian, Maximum Likelihood Estimation Binomial, Bayesian Gaussian, or Bayesian Beta-Binomial. Additionally, users have the flexibility to specify up to four racial/ethnic groups for comparison, choose a particular US state (or view nationwide predictions), and input specific values for available housing and sociodemographic variables such as high school completion rate (%), unemployment rate (%), annual household income (current US\$), and home value (current US\$). Furthermore, users can opt to display the predicted distributions separately or overlaid together.

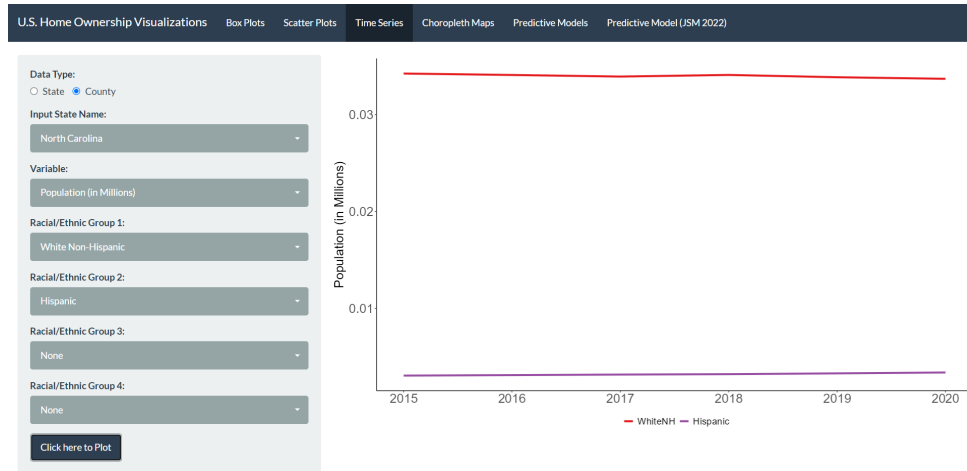


(a) Box Plots feature applied to county data in Nebraska for the year 2018. The violin plot feature was selected to visualize home ownership rates (%) and compare the White Non-Hispanic, Hispanic, Black, and Asian groups.

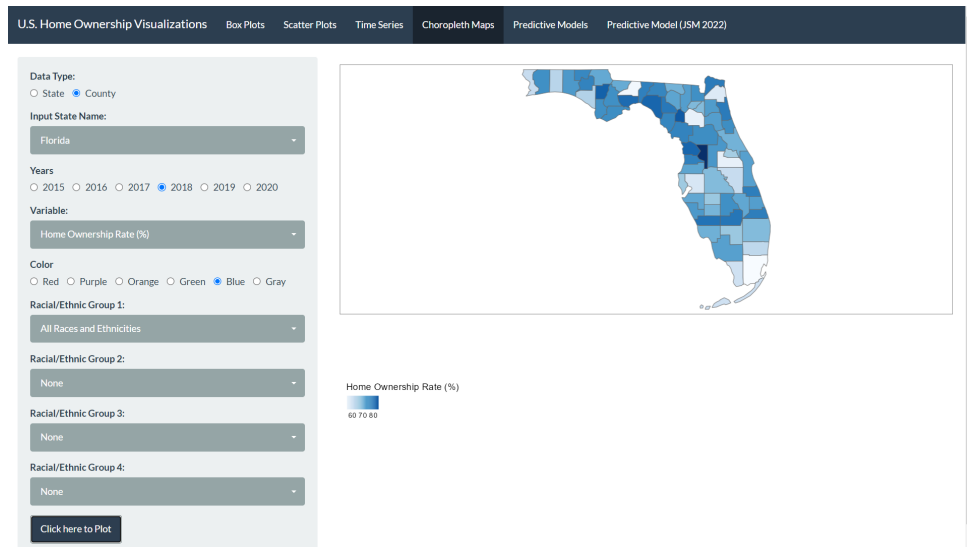


(b) Scatter Plots feature applied to county data in California for the year 2015. The locally estimated scatterplot smoother was used to visualize annual income (in current US\$ on a logarithmic base 10 scale) vs. high school completion (%) and compare the White Non-Hispanic, Hispanic, Black, and Asian groups.

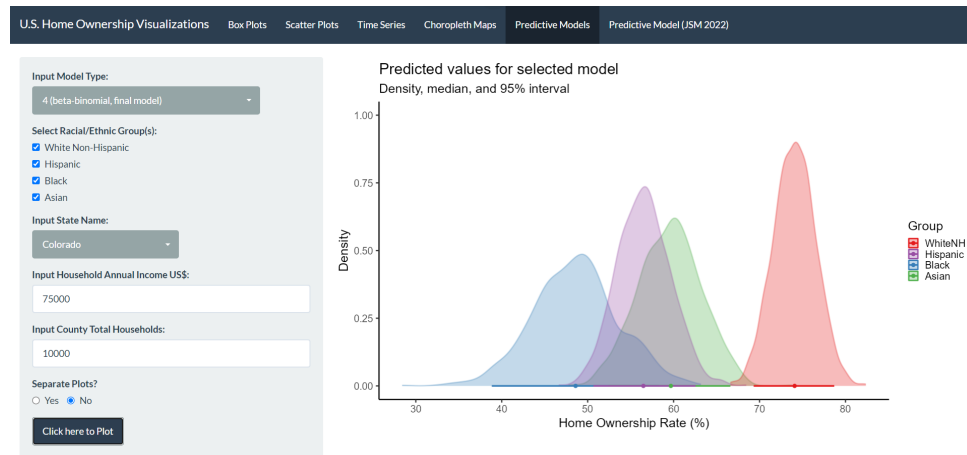
Figure S6: Shiny App Visualization Features



(c) Time Series feature applied to North Carolina county data. The population (in millions) variable was selected to compare the White Non-Hispanic and Hispanic groups over the period 2015-2020.



(d) Choropleth Maps feature applied to county data in the state of Florida for the year 2018. The home ownership rate (%) variable is visualized using a blue palette for all races and ethnicities.



(a) The plot illustrates the predicted home ownership rates for the White Non-Hispanic, Hispanic, Black, and Asian groups in the state of Colorado. The selected model is Model 4 (beta-binomial final model), and the plot showcases the predicted values based on an annual income of \$75,000 and county total households of 10,000. The four density curves are overlaid on the same baseline, although there is also an option to display them on separate baselines.



(b) The plot illustrates the predicted home ownership rates for the White, Black, and Asian groups in the state of California (JSM 2022 version). The selected model is the Bayesian Beta-Binomial, and the plot showcases the predicted values based on a state high school completion rate of 90%, unemployment rate of 5%, median annual household income of \$80,000, and median home value of \$200,000. The three density curves are overlaid on the same baseline, although there is also an option to display them on separate baselines.

Figure S7: Shiny App Predictive Features