2020 Census Predictive Models and Audience Segmentation Report

A New Design for the 21st Century

June 21, 2019
Version 1.5
Prepared by Robert Kulzick, Laura Kail, Shawnna Mullenax, Hubert Shang, Brian Kriz, Gina Walejko, Monica Vines, Nancy Bates, Steven Scheid, and Yazmín García Trejo

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

United States Census 2020

Table of Contents

List of Tables

List of Figures

Executive Summary

The goal of the 2020 Census Integrated Communications Campaign (ICC) is to encourage households to self-respond to the 2020 Census without follow-up from Census Bureau enumerators. The Communications Research and Analytics Team (CRAT) conducted the research described in this report to provide the U.S. Census Bureau and its communications contractor, Team Y&R, with research-based evidence that will help in identifying and delivering messages to the public in the manner and at the times and locations most effective for the 2020 Census communications campaign.

This report describes the use of statistical models and Census Bureau data sources to predict and understand patterns of self-response to the 2020 Census. It describes research on three aspects of self-response behavior: (1) predictions of the overall self-response rate, (2) the proportion of self-response submitted by internet for small geographic areas, and (3) response and self-response timing drawn from previous data collections at the national level, for small geographic areas, and for selected demographic groups.

The report then focuses on ICC segmentation, which sorts complex data related to self-response into two underlying typologies: (1) mindsets, which are combinations of knowledge about and attitudes toward the census that are shared by a group of people, and (2) tract segments—groups of census tracts with similar predicted self-response behavior and similar demographic variables associated with self-response—that were selected for their distinctive patterns of media consumption and distribution of mindsets.

The key findings from this study, organized by broad research question, are as follows:

What are the predicted self-response behaviors of census tracts to the 2020 Census?

- The 2020 Predicted Self-Response Score (2020-PSRS) predicts that census tracts, on average, will self-respond at a rate of 60.5 percent. The 2020-PSRS predicts that a quarter of U.S. households live in tracts that will self-respond at 51 percent or less and a quarter of U.S. households live in tracts that will self-respond at 70 percent or more.[1]
- The 2020 Internet Proportion of Self-Response (2020-IPSR) predicts that, on average, 66 percent of a tract's self-response will come through the internet. The 2020-IPSR predicts that a quarter of U.S. households live in tracts that will have 75 percent or less of their self-response online and a quarter of U.S. households will have 81 percent or more of their self-response online.

What do predicted 2020 Census self-response rates suggest about the self-response characteristics of designated market areas (DMAs)?

- The 10 DMAs with the highest predicted levels of self-response to the 2020 Census are located in the Upper Midwest and do not include the most populous U.S. cities.
- The 10 DMAs with the largest populations are predicted to self-respond at rates that are similar to the projected national self-response rate of 60.5 percent.

[1] Households in tracts without mailable addresses for the American Community Survey (ACS) were not given modeled scores and, therefore, were not included in this estimate.

- The 10 DMAs with the lowest predicted levels of self-response to the 2020 Census are in the Southeast and Southwest.

How might the timing of self-response vary across census tracts and demographic groups in the 2020 Census based on patterns in the timing of response and self-response to the 2010 Census and the timing of self-response to various census tests?

- Daily cumulative self-response rates to the 2010 Census at the tract level did not have abrupt changes over time. These rates had a smooth and consistent distribution across tracts for the entire period the Census Bureau reported them.
- Different demographic groups had different speeds of response to the 2010 Census. Older age groups responded more quickly than younger age groups. Among race and Hispanic-origin groups, non-Hispanic White responded most quickly, followed by non-Hispanic Asian; the remaining race and Hispanic-origin groups responded more slowly.
- The 2015 National Content Test (NCT) and the 2017 Census Test had consistent differences in self-response timing by response mode and contact strategy. Online submissions from households receiving an initial mailing with instructions for online self-response and no paper questionnaire (i.e. Internet First contact strategy) were the fastest form of self-response and made up the largest proportion of self-response for both tests.

What are the mindsets—that is, shared attitudes toward and knowledge of the census—present in the population, as reflected in the Census Barriers, Attitudes, and Motivators Study (CBAMS) Survey responses?

Six mindsets in the U.S. population were identified:

- The **Eager Engagers** mindset is comprised of the most civically engaged people who have the highest knowledge about the census. This group has the highest percentage of people who intend to respond to the census.
- The **Fence Sitters** mindset is the largest, making up 32 percent of the population. People with this mindset do not have major concerns about responding to the census.
- Individuals with a **Confidentiality Minded** mindset are the most concerned that their answers to census questions will be used against them, but they believe their answers matter.
- People with a **Head Nodders** mindset are the most likely to respond “yes” to all of the knowledge questions on the survey, and they have significant knowledge gaps in specific areas.
- The **Wary Skeptics** mindset is characterized by skepticism of the government and apathy about being counted in the census. Individuals with this mindset are reluctant to participate in the census.
- Individuals identified as having a **Disconnected Doubters** mindset do not use or have access to the internet, have above-average levels of apathy toward the census, and are least likely to respond to the census.

What are the distinguishing features of tract segments, or groupings of census tracts, in the United States? How are these tract segments geographically distributed?

There are eight tract segments in the U.S., and they range from highly likely to respond to the census to unlikely to respond. The eight tract segments are:

- **Responsive Suburbia** is predicted to self-respond at the highest rates overall and by internet. This segment is composed of census tracts that primarily consist of suburban neighborhoods of single-family homes.
- **Main Street Middle** is predicted to self-respond at a higher-than-average rate overall and by internet. This segment contains small towns and less densely populated neighborhoods surrounding urban centers.
- **Country Roads** is predicted to self-respond at an average rate overall with a below-average percentage of self-response coming online. This segment includes rural areas located predominantly in the eastern United States and small towns and areas outside the suburbs of major cities.
- **Downtown Dynamic** is predicted to self-respond at a slightly below-average rate with the majority of self-response coming online. This segment includes densely populated metropolitan centers.
- **Student and Military Communities** is predicted to self-respond at a slightly below-average rate with the majority of self-response coming online. This segment includes college campuses, military bases, and the towns that surround them. Only 2 percent of the U.S. population lives in the Student and Military Communities segment.
- **Sparse Spaces** is predicted to self-respond at a below-average rate with a below-average percentage of self-response coming online. It is composed of rural areas located predominantly in the western United States, Appalachia, northern Maine, and Michigan's Upper Peninsula.
- **Multicultural Mosaic** is predicted to self-respond at a below-average rate with a below-average percentage of self-response coming online. This segment has the highest proportion of people born outside the United States and the highest proportion of non-English-speaking households. A majority of people in this segment are Hispanic.
- **Rural Delta and Urban Enclaves** is predicted to self-respond at the lowest rate of all segments, with the lowest proportion of self-response coming online. A majority of this segment is non-Hispanic Black or African American.

1. Introduction

This report describes research that the 2020 Census Integrated Communications Campaign (ICC) will use, in conjunction with the 2020 CBAMS Survey and Focus Groups, to design an evidence-based campaign that will encourage participation in the decennial census (McGeeney et al., 2019; Evans et al., 2019). Specifically, communications professionals will use data on self-response and the ICC segmentation to make informed decisions about aspects of the campaign, including the allocation of spending for different communications channels (e.g., TV, radio, digital), advertising messages and content, and the timing of spending during the campaign. These strategic choices will involve professionals from numerous communications disciplines and will occur across many independent decision points. For a detailed description of the communications activities and strategies of the ICC, see the 2020 Census Integrated Communications Plan.

To plan a campaign that encourages self-response, the ICC will use predictions about self-response within small geographic areas to identify and devote additional resources to areas with relatively lower predicted self-response. These predictions include the overall rate of self-response, as well as the proportion of self-response that occurs online. Predictions for these small geographic areas allow communications professionals to predict the self-response rates of larger geographic areas by aggregating the predictions of their component parts.

In addition to predictions about self-response, the ICC will use information about the timing of response and self-response drawn from previous Census Bureau data collections. The timing of self-response is particularly relevant to planning the timing of advertising and assessing the performance of the communications campaign during the enumeration. For communications directed to a particular demographic group, the speed of that group's response during past data collections will inform the amount of resources allocated to later phases of the campaign.

Finally, the ICC will use one typology of individuals and one typology of small geographic areas to divide the country into distinct groupings that highlight demographic and attitudinal characteristics relevant to the communications campaign. Mindsets are sets of knowledge of and attitudes toward the 2020 Census shared by many individuals. Team Y&R will use mindsets to create messages that address the full range of attitudes about and concerns with the 2020 Census. Tract segments identify census tracts that share common predicted self-response and demographic features related to self-response. Team Y&R will use tract segments to divide the country into regions that can receive specialized messaging mixes and amounts of communications that reflect their predicted levels of self-response. Mindsets and tract segments characterize a diverse nation in a simple manner while retaining the most distinctive and relevant aspects of that diversity.

2. Background

2.1 Predicting Self-Response Using the Low Response Score (LRS)

The U.S. Census Bureau has been conducting research into factors influencing survey response rates since at least the 1990s. To prepare for the 2010 Census, researchers at the Census Bureau developed a hard-to-count (HTC) score to identify areas that might present enumeration challenges (Bruce et al., 2001). The HTC score was calculated for each census tract based on data from the 2000 Census and a set of 12 variables that had been shown in previous research to be correlated with high rates of nonresponse to Census Bureau surveys. The 12 variables were equally weighted in HTC score calculations. The Census Bureau used the HTC score in planning the 2010 Census and other surveys conducted by the Census Bureau.

After the 2010 Census, the Census Bureau sought to develop a version of the HTC score based on statistical methods that directly model census tract self-response rates. In 2012, the Census Bureau posed the Kaggle “Census Return Rate Challenge” to data scientists; the challenge was to predict 2010 Census mail return rates using the 2012 Census Planning Database (PDB) and any other publicly available sources of data. The winners used ensemble machine learning techniques that combined Census Bureau data with publicly available data to model self-response.

In 2014, the Census Bureau used the findings from the Kaggle challenge to introduce a major update to the HTC score, the low response score (LRS) (Erdman and Bates, 2017). The LRS modeled the 2010 Census mail nonreturn rate and, therefore, has higher values in tracts with lower 2010 Census mail return rates—that is, the higher the LRS score, the lower the mail return rate. To ensure this score would be actionable and transparent, the Census Bureau did not directly use the winning algorithm from the Kaggle challenge; instead, the Census Bureau used an ordinary least squares regression approach based solely on a group of PDB variables that had been found to be highly predictive in the winning challenge model. Since 2014, the Census Bureau has updated the LRS score with each new version of the PDB.

The CRAT team is building on this research to create a predicted self-response score (PSRS) to reflect the next decennial census environment. As opposed to the LRS, which predicts the proportion of households that will *not* self-respond, the PSRS predicts the proportion that *will*.

The introduction of an internet response option is one of the major changes to the 2020 Census design, so rather than model the PSRS using mail response to the 2010 Census, which did not have an internet self-response option, the CRAT team used self-response to the American Community Survey (ACS), which has had an internet response option since 2013. The CRAT team will develop an ACS-PSRS and ACS Internet Predicted Self-Response Score (ACS-IPSRS) that reflect the current ACS environment and modes of response. These models will serve as a starting point for the Integrated Partnership and Communications (IPC) team to create a 2020-PSRS and 2020 Internet Proportion of Self-Response (2020-IPSR) that predict overall self-response to the 2020 Census and the proportion of that self-response that will be online.

2.2 Adjusting ACS Predictions to Reflect the 2020 Census Environment

The 2020 Census will offer new contact strategies, Internet First and Internet Choice, related to the introduction of a widely available internet response option.

The Internet First contact strategy has an initial mailing that includes instructions on how to complete the survey through the internet instrument but does not include a paper questionnaire in the first mailing. The Census Bureau plans for about 80 percent of housing units in the mailing universe to receive the Internet First contact strategy.

The 20 percent remaining—those housing units identified as less likely to respond by internet and with lower rates of self-response—will also have the option to respond immediately by paper. Called Internet Choice, this alternative contact strategy sends a paper questionnaire along with information on how to respond online in the first mailing (U.S. Census, 2018), giving respondents a choice between internet and mailed questionnaire modes of response.

Tests conducted by the Census Bureau in preparation for the 2020 Census show large differences in the amount of online self-response these two contact strategies elicit. As an example, the Internet Choice panel in the 2015 National Content Test (NCT) had an internet proportion of self-response of 25 percent, and the Internet First panel had a much higher internet proportion of self-response, of 70 percent (Letourneau, 2018). The ACS has used a contact strategy similar to Internet First since the introduction of an online self-response option. However, the proportion of online self-response to the ACS in 2013 was around 55 percent (Baumgardner, 2014). This lower proportion of online response compared with the Internet First panel in the 2015 NCT, as well as the absence of a contact strategy equivalent to Internet Choice, means that a dual adjustment that is dependent on contact strategy is necessary when using ACS-based predictions to produce the 2020-IPSR.

Projected self-response rates will also have to be adjusted for differences between projected overall 2020 Census response rates and overall ACS self-response rates. Self-response to the ACS is referred to as self-check-in,[2] and the average national self-check-in rate for both 2012 and 2013 is close to 50 percent (Baumgardner, 2014), while the national projected self-response rate for the 2020 Census is 60.5 percent (U.S. Census, 2017).

[2] When this report refers to or discusses data from the ACS, the phrase "self-check-in" denotes self-response behavior. Self-response is an unassisted response to a survey or census form that does not include telephone questionnaire assistance responses, in-person responses captured through computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), or the assistance of an in-person enumerator. Self-check-in, as defined by the ACS, is the total number of ACS internet returns before the start of CATI and mail returns up to seven days after the start of CATI from mailable addresses, divided by the total number of mailable housing unit addresses sampled in the tract (excluding group quarters).

3. Creating the 2020-PSRS and 2020-IPSR

3.1 Methodology

The first step in predicting variation in self-response to the 2020 Census is modeling both overall self-response and internet self-response behavior to the ACS. The second is adjusting those scores to reflect the decennial environment.

3.1.1 ACS Modeling

This section provides a high-level narrative of the steps taken to produce ACS models, which then predict overall self-check-in rates and internet self-check-in rates at the tract level.

3.1.1.1 Data Sources

The CRAT team began by calculating the ACS overall self-check-in rate and internet self-check-in rate at the tract level from ACS paradata at the housing unit level. Because the Census Bureau added an internet response mode to the ACS in 2013, the CRAT team calculated all self-check-in rates within the 2013-2016 period. To produce self-check-in rates over periods, the CRAT team pooled responses from the same tract over the entire period. These served as the dependent variables for all modeling efforts. Technical Appendix B.1.1 includes additional details on how the CRAT team calculated self-check-in rates.

$$\text{Overall Self-Check-In Rate}_{\text{tract}} = \frac{\sum_i^{n_{tract}} \text{ACS Base Weight}_i \times \text{Mail or Internet Self-Response Flag}_i}{\sum_i^{n_{tract}} \text{ACS Base Weight}_i} \quad (1)$$

$$\text{Internet Self-Check-In Rate}_{\text{tract}} = \frac{\sum_i^{n_{tract}} \text{ACS Base Weight}_i \times \text{Internet Self-Response Flag}_i}{\sum_i^{n_{tract}} \text{ACS Base Weight}_i} \quad (2)$$

The CRAT team used two overlapping sets of independent variables—a smaller set of core variables and a larger set of expanded variables—to predict overall and internet ACS self-check-in rates. The expanded set of variables included all of the tract-level variables available in the PDB and their appropriate transformations. Technical Appendix B.1.2.2 contains details on the expanded variables.

The core variables were composed of demographic variables identified in the original 2014 LRS study and other variables predictive of ACS self-check-in rates by Census Bureau researchers. Research by the Census Bureau's Center for Statistical Research and Methodology (CSRM) also identified four PDB variables useful for predicting ACS self-check-in rates but not included in the original 2014 LRS study. The CRAT team included these variables in its analysis: the percentage of a census tract that (1) speaks a language other than English at home, (2) lives in a house with more than one person per room, (3) lacks phone service, and (4) lives in housing units considered mobile homes.[3] Table 9 in Technical Appendix B.1.2 provides a detailed list of all PDB variables included in the list of core variables.

In addition to the PDB variables, the CRAT team included, as a categorical variable, the Esri Tapestry[4] segments because research by the Census Bureau has shown that the Esri Tapestry Segmentation assignments predict ACS self-check-in rates (Mulry, Bates, and Virgile, 2018). The CRAT team also included three additional core variables that CSRM identified as effective in predicting ACS self-check-in rates (see Table 10 in Technical Appendix B.1.2 for a detailed list of all non-PDB variables included in the list of core variables). The first is a measure of urban density produced by the Census Bureau's Geography Division. This tract-level urban density variable measures the proportion of land area taken up by urban blocks in a census tract. The CRAT team developed the two other core variables from questions added to the ACS in 2013 that measured internet access. Variables derived from responses to the internet access questions (see Table 11 in Technical Appendix B.1.2) produced a measure of internet access and a measure of nonmobile internet access.[5] After creating these transformed variables for each sampled household from 2013 through 2016, the CRAT team weighted each household using its single-year base weight, aggregated weighted self-responses into single-year tract-level estimates for each year, and averaged across the four years of available self-response data.

3.1.1.2 Model Estimation and Validation

The CRAT team estimated four types of models to determine the best way to predict overall and internet ACS self-check-in rates with the goal of choosing one model. These models included two linear models—ordinary least squares (OLS) regression and least absolute shrinkage and selection operator (LASSO) regression—and two decision tree models, random forest and gradient-boosted trees. The LASSO regression and both decision tree models discard unhelpful independent variables. Because of this, the LASSO, random forest, and gradient-boosted tree models used the expanded set of variables, and the OLS models used only the core variables. Technical Appendix B.1.3 contains details on these models' algorithms.

To reduce the influence of sparsely populated tracts, the CRAT team assigned each tract a weight equivalent to the number of households within it. Using these weights in estimating our models made information from tracts with more housing units more influential in determining model parameters.

---

[3] The percentage of a census tract that speaks a language other than English at home is calculated as the proportion of the population five years or older in the tract. The percentage of a census tract that lives in a house with more than one person per room and the percentage of a census tract that lacks phone services are calculated as the proportion of the total number of occupied housing units in the tract. The percentage of a census tract that lives in housing units considered mobile homes is calculated as the proportion of total housing units in the tract.
[4] The Esri Tapestry Segmentation is a market segmentation database that groups neighborhoods with similar psychographic, social, and demographic characteristics. Esri bases these segments on an aggregation of consumer marketing, census responses, and other lifestyle data sets.
[5] The percentage of a census tract that has internet access and the percentage of a census tract that has nonmobile internet access are calculated as the proportion of the total number of occupied housing units in the tract.

To select the final ACS-PSRS and ACS-IPSRS models, the CRAT team compared the candidate models to determine which algorithm produced the best out-of-sample fit—in other words, which was the most generalizable. The CRAT team used the single-year overall and internet self-check-in rate from 2016 as a validation data set. Using this system, the model that best predicted the 2016 self-check-in rates using past tract-level characteristics would be the best choice for predicting future self-check-in rates using current tract-level characteristics.

After developing each model from the training data set, the CRAT team applied it to the validation data set to make predictions for model evaluation. A single metric gave the primary means of comparison: the weighted mean absolute error (WMAE) between the actual 2016 ACS self-check-in rate and the predicted 2016 ACS self-check-in rate for each modeling approach. Technical Appendix B.1.4 includes additional details on the model validation approach.

$$\text{WMAE} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i |\hat{y}_i - y_i| \qquad (3)$$

where w_i is the number of housing units within the tract

3.1.1.3 ACS Model Selection

The CRAT team selected the OLS regression as the best modeling approach for predicting ACS overall and online self-check-in rates. In out-of-sample testing, OLS models did not produce candidate models with the lowest out-of-sample WMAE. However, the reduced operational complexity and greater interpretability of the OLS models justified choosing them despite a small loss in predictive accuracy compared with the best fitting alternative models. See Table 13 in Technical Appendix B.1.5 for error metrics used to assess the various model candidates.

3.1.1.4 Disclosure Review Board Noise Injection

The CRAT team developed the ACS-PSRS and ACS-IPSRS models using data that were, for the most part, publicly available. However, ACS self-check-in rates were not publicly available at the time. Therefore, creating the dependent variables for both the ACS-PSRS and ACS-IPSRS models required the use of ACS operations data, protected by Title 13, to calculate the overall self-check-in rate, as well as the internet self-check-in rate for years 2013 through 2016.

Similarly, the ACS began asking about internet access in 2013, but did not provide publicly available estimates of internet access while the CRAT team's modeling activities were occurring. Therefore, to create internet access variables for use as an independent variable, the CRAT team used ACS responses protected by Title 13.

To ensure models produced using Title 13 data can be disclosed without violating Title 13, the CRAT team used a process of noise injection consistent with the principles of differential privacy on all non-publicly available variables. For noise injection, the CRAT team used the Laplace Mechanism, which adds random values from a Laplace distribution with mean zero and a scale parameter of $\frac{\Delta f}{\varepsilon}$, where Δf is the sensitivity of the query and ε is a privacy parameter.[6] See Technical Appendix B for additional disclosure avoidance details.

3.1.1.5 Scoring and Final Estimation

After deciding on the appropriate estimation technique for calculating the ACS-PSRS and ACS-IPSRS and after performing the noise injection required for disclosure avoidance, the CRAT team re-estimated the chosen models using the ACS self-check-in rates from 2013 through 2016. This provided the largest possible sample sizes in each tract and ensured that all available self-check-in rate data was used to make our final predictions.

3.1.2 Decennial Adjustments

3.1.2.1 Adjusting the ACS-PSRS

To reflect self-response rates appropriate for a decennial census, the CRAT team adjusted the ACS-PSRS to the 2020-PSRS by adding a constant to each tract score that ensured the tract average weighted by the number of households would match the national projected 2020 self-response rate of 60.5 percent (U.S. Census, 2017). The CRAT team calculated self-response rates following the definition of response rates rather than return rates because information on deletes, vacants, and undeliverable as addressed households will not be publicly available during the 2020 Census, meaning that the communications campaign will have to rely on response rates. The overall ACS self-check-in rate for tracts was calculated by pooling self-response from 2013 through 2016. The CRAT team then aggregated tract-level ACS rates to the national level using a population-weighted average of tracts, as detailed in the following equations:

$$\text{2020-PSRS}_{tract} = \text{National 2020 Projected Self-Response Rate} - \text{Household-Weighted Average ACS-PSRS} + \text{ACS-PSRS}_{tract} \quad (4)$$

$$\text{Household-Weighted Average ACS-PSRS} = \frac{\sum_i^n \text{Total Households}_{tract_i} \times \text{ACS-PSRS}_{tract_i}}{\sum_i^n \text{Total Households}_{tract_i}} \quad (5)$$

3.1.2.2 Adjusting the ACS Internet Proportion of Self Response (ACS-IPSR)

Next, the CRAT team adjusted the ratio of ACS-Internet Predicted Self-Response Score (IPSRS) and ACS-Predicted Self-Response Score (PSRS) or the ACS Internet Proportion of Self-Response (IPSR) separately for the Internet First and Internet Choice contact strategies. This research used the most recent contact strategy assignments available.[7]

---

[6] Dwork and Roth (2014) describe the Laplace Mechanism and specific definitions of query sensitivity.

[7] Census Bureau researchers provided the preliminary contact strategy file on April 11, 2018.

To calculate adjustments for each contact strategy, the CRAT team first placed data from both the 2015 NCT and 2017 Census Test into two separate data sets based on contact strategy[8] and then calculated the internet proportion of self-response for each tract. Next, the population-weighted average internet proportion of self-response was calculated for each contact strategy. This calculation included two separate estimates from any tracts that were in-sample for both tests. The adjustment for each contact strategy is the difference between the weighted average internet proportion of self-response of all tracts with that contact strategy in the 2015 NCT and 2017 Census Test and the population-weighted average of the ACS-IPSR.

$$\text{Internet First Adjustment} = \frac{\sum_i^{n_{\text{Internet First}}} \text{Population}_{tract_i} \times \text{IPSR 2015 NCT and 2017 CT}_{tract_i}}{\sum_i^{n_{\text{Internet First}}} \text{Population}_{tract_i}} - \frac{\sum_i^{n} \text{Population}_{tract_i} \times \text{ACS-IPSR}_{tract_i}}{\sum_i^{n} \text{Population}_{tract_i}} \quad (6)$$

$$\text{Internet Choice Adjustment} = \frac{\sum_i^{n_{\text{Internet Choice}}} \text{Population}_{tract_i} \times \text{IPSR 2015 NCT and 2017 CT}_{tract_i}}{\sum_i^{n_{\text{Internet Choice}}} \text{Population}_{tract_i}} - \frac{\sum_i^{n} \text{Population}_{tract_i} \times \text{ACS-IPSR}_{tract_i}}{\sum_i^{n} \text{Population}_{tract_i}} \quad (7)$$

Because these adjustments were applied to logit-transformed ACS-IPSR scores, the CRAT team applied a logit transformation to the adjustments. Then, an inverse logit transformation was used on the sum of the logit-transformed ACS-IPSR and the logit-transformed adjustment to produce a 2020-IPSR between zero and one. Technical Appendix B.2 describes additional details of the ACS-IPSR adjustment.

For tracts with Internet First contact strategy in 2020:

$$\text{2020-IPSR}_{tract} = \text{ACS-IPSR}_{tract} + \text{Internet First Adjustment} \quad (8)$$

For tracts with Internet Choice contact strategy in 2020:

$$\text{2020-IPSR}_{tract} = \text{ACS-IPSR}_{tract} + \text{Internet Choice Adjustment} \quad (9)$$

[8] The CRAT team considered six panels out of nine from the 2015 NCT that are slight variations on the Internet First contact strategy as receiving the Internet First contact strategy. Two panels that were not included in calculations of the adjustments for the ACS-IPSR had more significant deviations from the control Internet First contact strategy. Panel 4 received a paper questionnaire two weeks earlier than is typical of an Internet First contact strategy, and Panel 8 never received a paper questionnaire. The remaining panel received the Internet Choice contact strategy (Phelan, 2016).

3.2 Results

3.2.1 Self-Response Predictions by Tract

3.2.1.1 Statistical Distribution

At the tract level, the 2020-PSRS predicts that, on average, tracts will self-respond at a rate of 60.5 percent. It also predicts that 25 percent of U.S. households will be in tracts that self-respond to the 2020 Census at a rate below 51 percent, and 25 percent of U.S. households will be in tracts that self-respond to the 2020 Census at a rate above 70 percent (see Table 1). The 2020-IPSR predicts that, on average, tracts will have 66 percent of their self-response through the internet, and 75 percent of U.S. households will be in tracts that have more than 64 percent of their self-response online.[9]

Table 1. Tract-Level Summary Statistics Weighted by the Number of Households

Statistic	2020-PSRS	2020-IPSR
Mean	60.5%	66%
Standard Deviation	13%	25%
25th Percentile	51%	64%
Median	62%	76%
75th Percentile	70%	82%

The distribution of the 2020-PSRS shows a single-peaked distribution with a thick left tail (see Figure 1). It has a wide, rather than a narrow, peak and shows a larger concentration of households in tracts with lower 2020-PSRS values than a normal distribution would produce.

[9] Households in tracts without mailable addresses for the ACS were not given modeled scores and, therefore, were not included in this estimate.

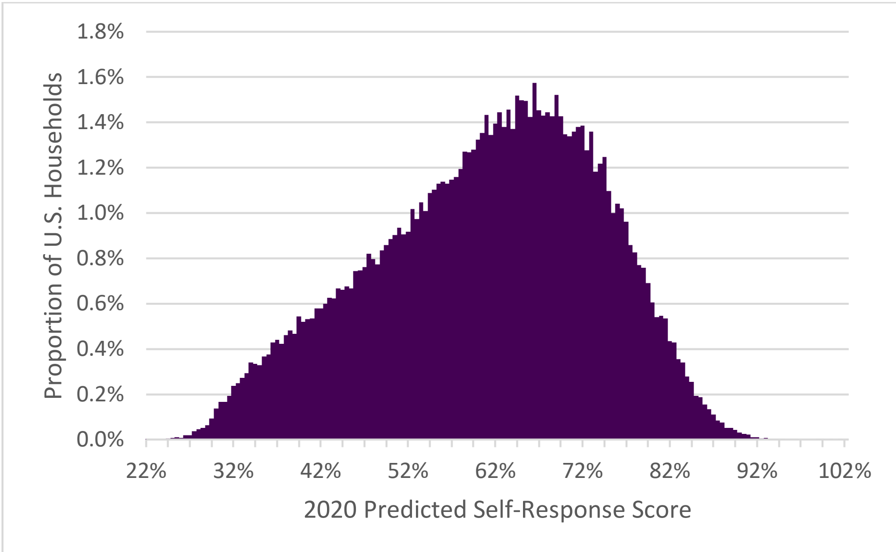


Figure 1. Weighted Distribution of the 2020-PSRS

The distribution of the 2020-IPSR is bimodal with separate distributions for Internet First and Internet Choice contact strategies caused by the adjustment described in Section 3.1.2.2 (see Figure 2). These two distributions share a similar width. The fact that there are more households with the Internet First distribution results in a much higher peak.

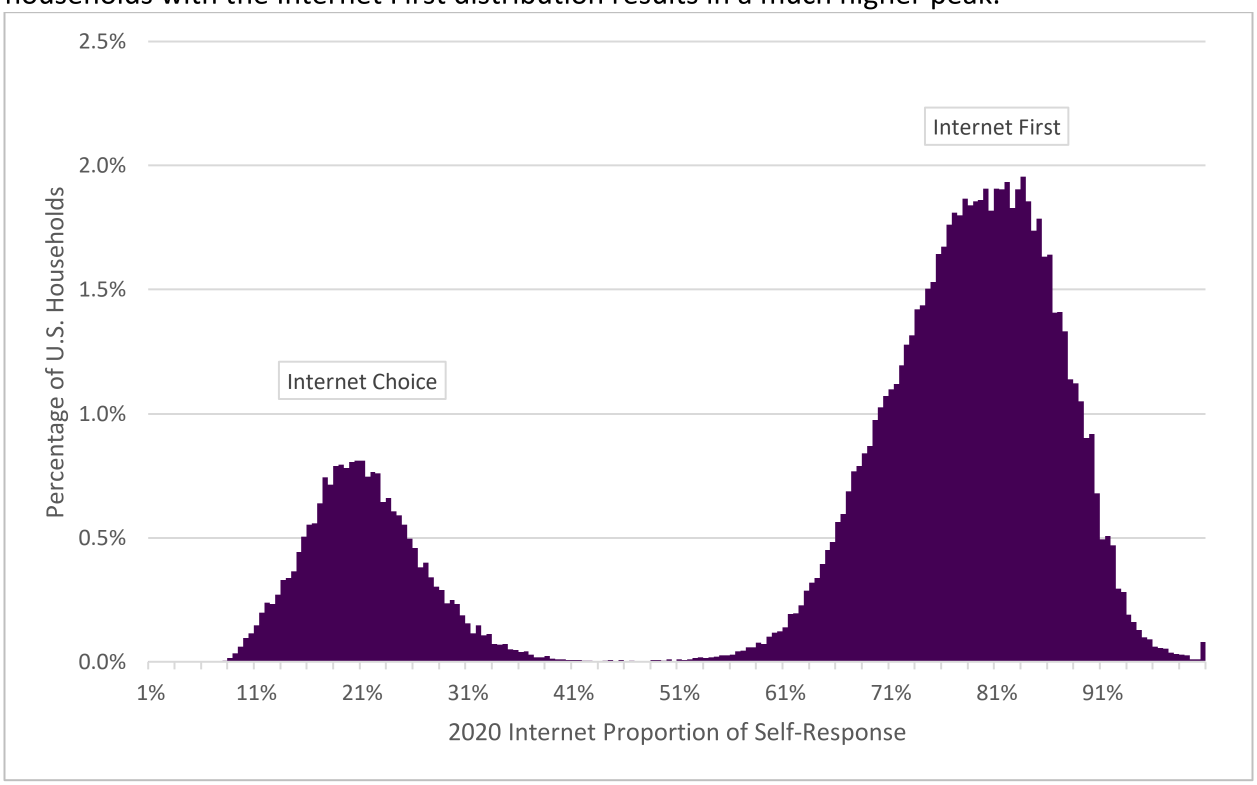


Figure 2. Weighted Distribution of the 2020-IPSR

3.2.1.2 Geographic Distribution

The national view of the 2020-PSRS (Figure 3) shows that tracts with higher 2020-PSRS values are concentrated in the Upper Midwest and Northeast corridor. In contrast, areas of low 2020-PSRS values exist in portions of the southern United States, as well as in the southwestern and western United States. Areas in gray did not participate in the ACS through the mail, and therefore do not have predicted 2020-PSRS scores.

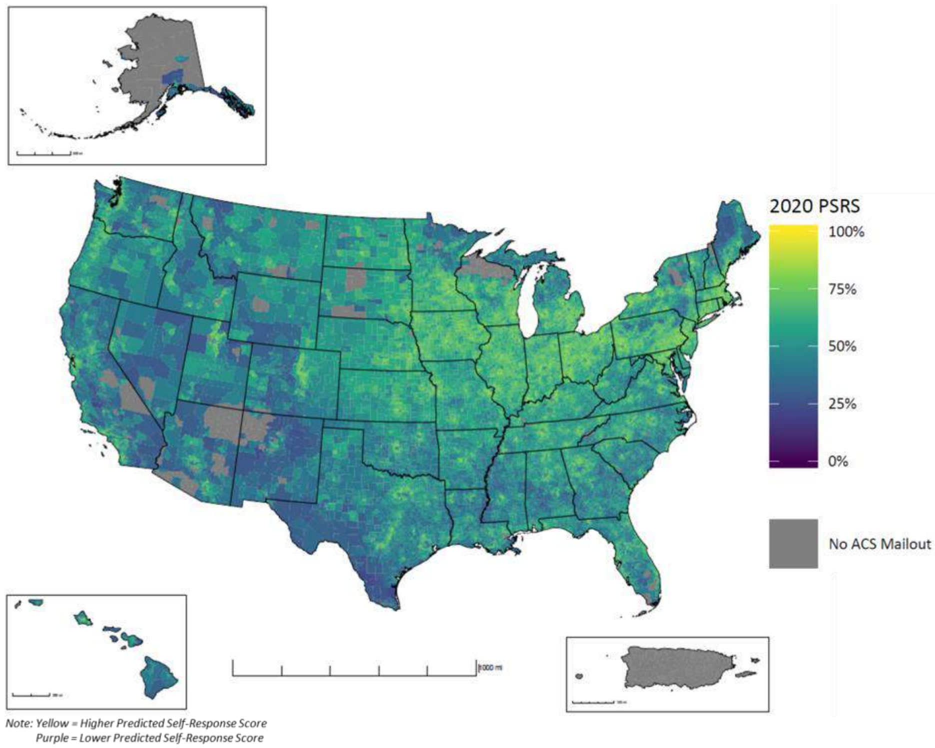


Figure 3. National View of the 2020-PSRS

The Washington, D.C., map of the 2020-PSRS (Figure 4, below) shows some areas with high predicted rates of self-response in northern D.C., but, in general, shows rates around 75 percent in many areas of Northwest D.C. In the northeastern and southeastern portions of D.C., the rate of predicted self-response is lower, with many of the lowest values concentrated in Southeast D.C.

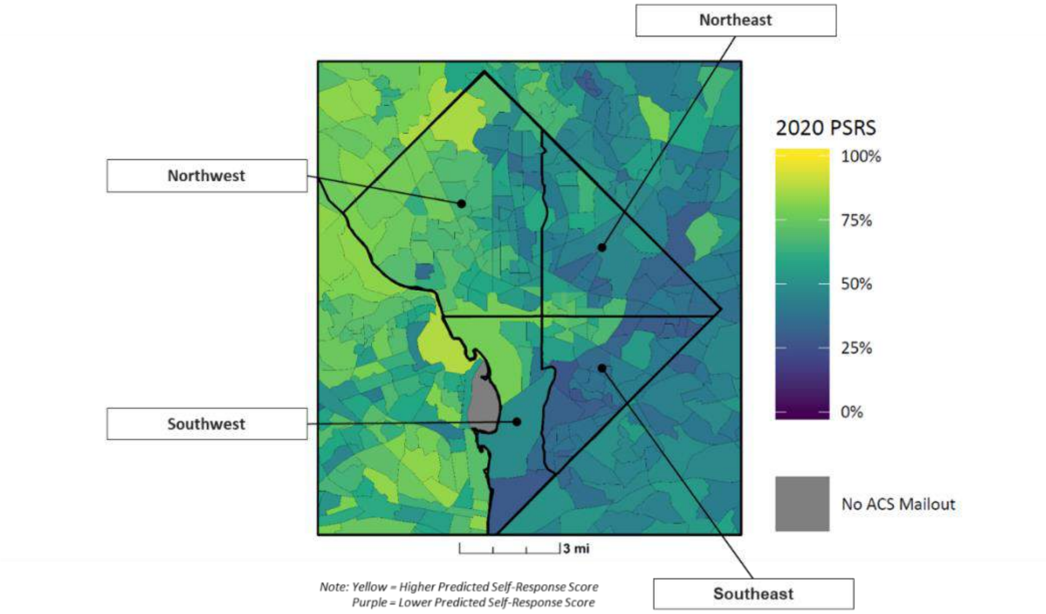


Figure 4. Washington, D.C. View of the 2020-PSRS

Because a tract's 2020-IPSR value depends on its assigned contact strategy, the distribution of 2020-IPSR values reflects the distribution of contact strategy. The national view of the 2020-IPSR in

Figure 5 shows that areas predicted to have high proportions of online self-response are distributed across much of the country, reflecting the fact that the Census Bureau has assigned the Internet First contact strategy to 80 percent of the population. Areas that show concentrated low proportions of online self-response are easier to isolate. These areas will receive the Internet Choice contact strategy. Northern Maine, northern Michigan, areas around the southern half of the Mississippi River, regions in northern California and Oregon, southern Texas, and scattered areas throughout the Midwest are all predicted to have a low proportion of their self-response come via the internet.

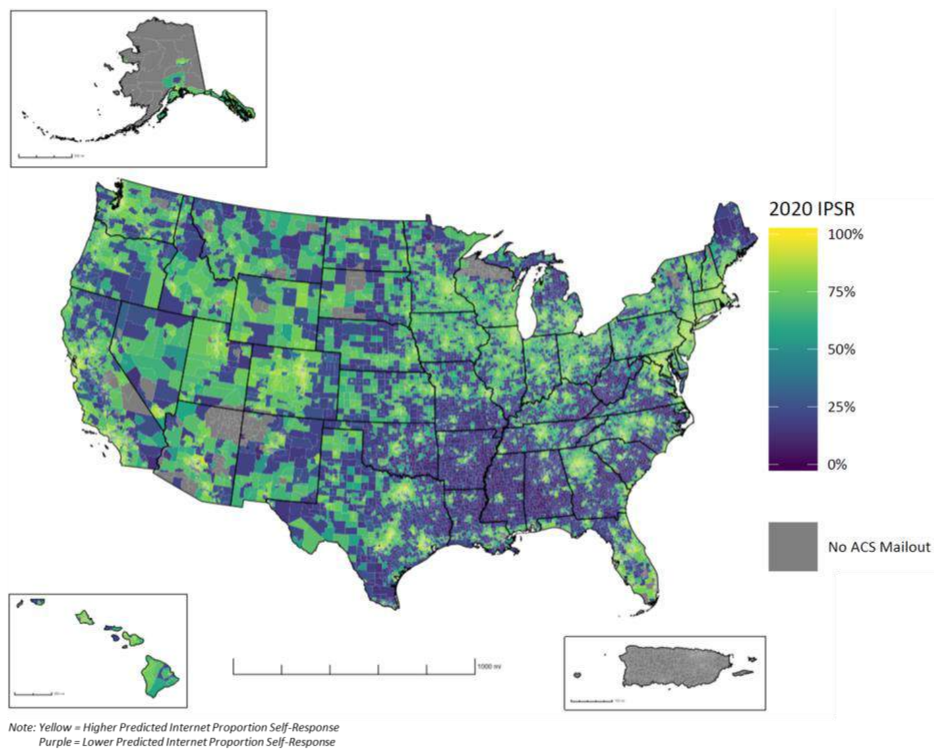


Figure 5. National View of the 2020-IPSR

In Figure 6, a map of the distribution of 2020-IPSR scores in Washington, D.C., shows fairly high predicted proportions of online self-response for Northwest D.C., and concentrations of low predicted proportions of online self-response in southern and Southeast D.C. Northeast D.C. shows some areas with a low predicted proportion of online self-response and areas that are in the middle.

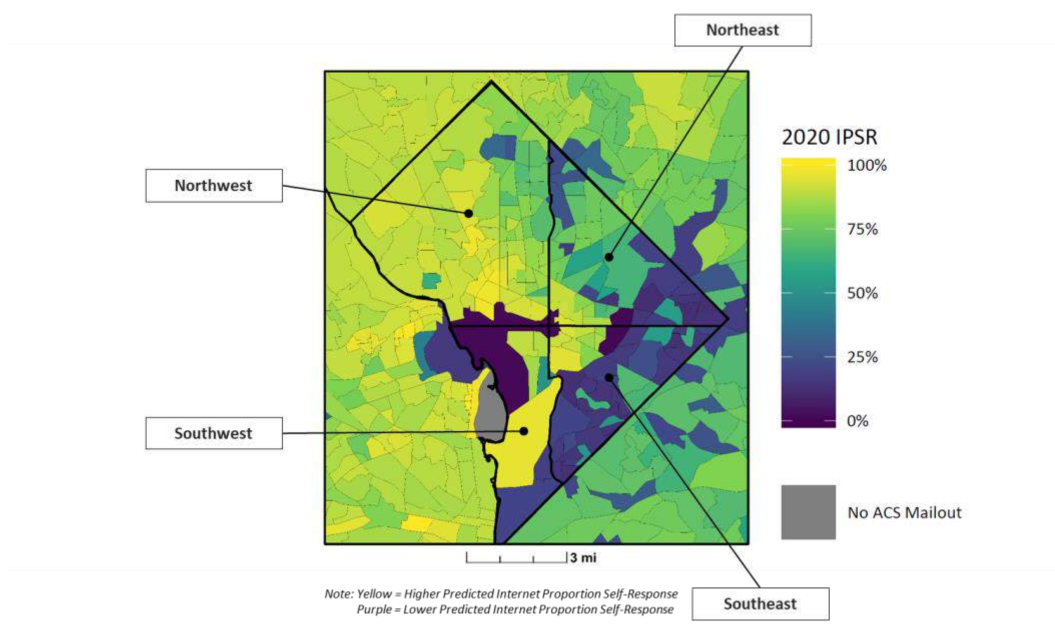


Figure 6. Washington D.C., View of the 2020-IPSR

3.2.2 Self-Response Predictions by DMA

3.2.2.1 Understanding Self-Response Predictions

Census tracts are relatively small geographic areas, from which it can be difficult to derive an overall impression of the predicted self-response score (2020-PSRS) and percentage of self-response by internet (2020-IPSR). However, these tract-level predictions allow the communications campaign to flexibly aggregate predictions to larger geographic areas that are more meaningful for the communications campaign, like a designated market area (DMA).

DMAs divide the country into geographic media markets used by those purchasing local television advertising. They represent an important geographic unit because local television has the largest reach among media channels that allow messages customized to a particular geography. With a few minor exceptions, DMAs contain entire counties, which makes it simple to produce DMA-level aggregates using the weighted average of tracts within the DMA.[10] Because our models predict self-response, which occurs at the household level, the number of households[11] in each tract was used as the weighting variable when producing DMA-level predictions. Similar procedures can produce predictions for other geographies. The only aggregate geography included in this report is the DMA, because of its particular importance and because it serves as an example of the aggregation of tract-level predictions to larger geographies.

Aggregating tract-level predictions reduces the amount of variation in our predictions. As seen in Figure 7, each modeled score has a smaller interquartile range and less extreme minimum and maximum values when aggregated to the DMA.[12] For the 2020-IPSR, the tract-level distribution has a wider distribution of values below the median. And when the data are aggregated to the DMA level, the fact that the area between the median and the first quartile is larger than that between the median and the third quartile shows the presence of DMAs with many Internet Choice contact strategy tracts.

---

[10] In cases where counties are in multiple DMAs, the CRAT team assumed an even distribution of self-response and divided the number of responses between DMAs using the proportion of households in the county that are in each DMA.

[11] Household estimates are taken from the ACS Total Occupied Housing Units estimates from the 2018 PDB.

[12] The maximum value of the 2020-PSRS at the tract-level exceeds 100 percent because the adjustment of the ACS-PSRS moved some tracts with high values of ACS-PSRS above 100 percent. While self-response rates above 100 percent are not possible, these values were maintained to capture the high confidence of our prediction that these tracts will respond at a very high rate.

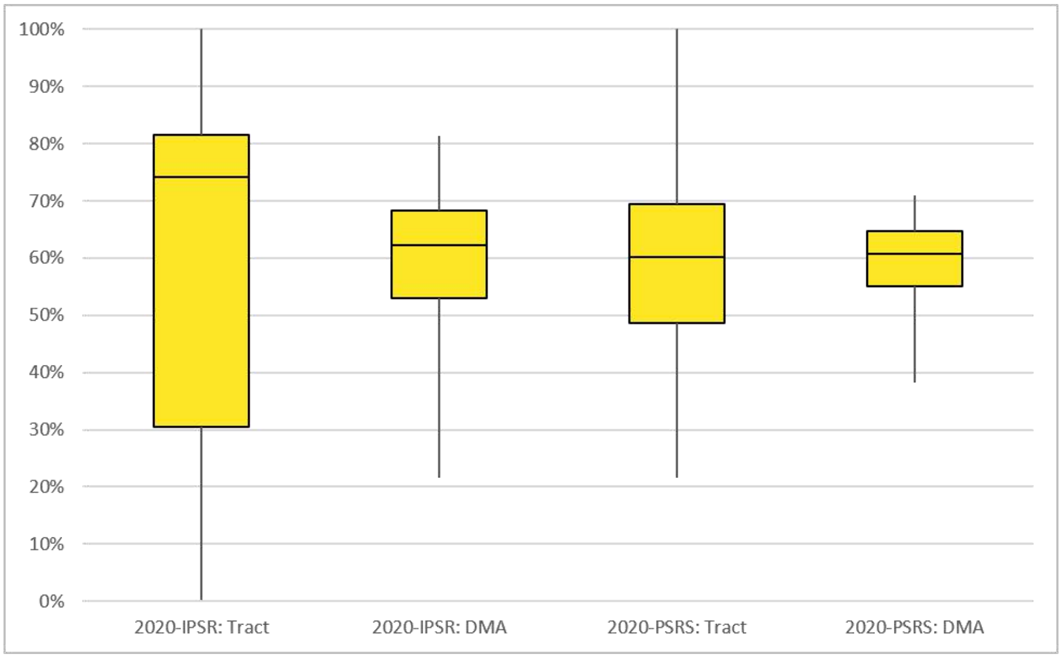


Figure 7. Distribution of Unweighted Modeled Scores at the Tract and DMA Level

3.2.2.2 Self-Response Predictions by DMA

Table 2 shows the 10 DMAs with the highest predicted levels of self-response to the 2020 Census. These DMAs are located in the Upper Midwest and, for the most part, do not include major cities. The 2020-IPSR predicts that most DMAs will have a proportion of online self-response that is similar to their predicted self-response rate. The exception in this list is the Parkersburg DMA (in Ohio and West Virginia), which has an internet proportion of self-response that is lower than the other top-ranked DMAs. The predicted self-response rate for these DMAs tends to be close to 70 percent, which is almost 10 percentage points higher than the projected national self-response rate of 60.5 percent.

Table 2. The 10 DMAs With the Highest Predicted Self-Response to the 2020 Census

DMA	State(s)	2020-PSRS	2020-IPSR
Cedar Rapids-Waterloo-Iowa City and Dubuque	Iowa	71%	70%
Lima	Ohio	70%	64%
Rochester-Mason City-Austin	Iowa, Minn.	70%	70%
Harrisburg-Lancaster-Lebanon-York	Pa.	70%	69%
Parkersburg	Ohio, W.Va.	69%	55%
Pittsburgh	Md., Pa., W.Va	69%	64%
Peoria-Bloomington	Ill.	69%	68%
Mankato	Minn.	69%	68%
Des Moines-Ames	Iowa	69%	69%
Minneapolis-St. Paul	Minn., Wis.	68%	75%

In addition to considering the DMAs with the highest rates of self-response, it is important to consider the amount of self-response that is predicted to come from the largest DMAs. Table 3 shows the largest DMAs in the United States in order of number of TV homes (Nielsen, 2019). For the most part, these DMAs have predicted rates of self-response within a point or two of the projected national response rate of 60.5 percent. Los Angeles and Houston are somewhat lower than the other large DMAs, with predicted self-response rates around 55 percent, and Boston and Philadelphia are higher. The predicted proportion of online self-response is similar across DMAs, with San Francisco-Oakland-San Jose, Washington, D.C., and Boston having predicted rates of online self-response that are around 6 percentage points higher than the other areas.

Table 3. The 10 DMAs With the Most TV Homes and Their Predicted Self-Response to the 2020 Census

DMA	State(s)	2020-PSRS	2020-IPSR
New York	Conn., NJ, NY, Pa.	58%	69%
Los Angeles	Calif., Nev.	55%	71%
Chicago	Ill., Ind.	61%	71%
Philadelphia	Del., NJ, Pa.	65%	68%
Dallas-Fort Worth	Texas	59%	71%
Washington, D.C. (Hagerstown)	DC, Md., Pa., Va., W.Va.	63%	79%
Houston	Texas	55%	69%
San Francisco-Oakland-San Jose	Calif.	62%	80%
Boston (Manchester)	Mass., NH, Vt.	67%	77%
Atlanta	Ala., Geo, NC	58%	73%

The CRAT team predicts some DMAs will be challenging from a self-response perspective. Table 4, below, shows the 10 DMAs with the lowest predicted rates of self-response to the 2020 Census. Geographically, these low-response DMAs are in the southeastern or southwestern United States. The two lowest-responding DMAs have predicted self-response rates of around 38 percent—noticeably lower compared with other DMAs in this group.

There is an interesting range of predicted proportion of online self-response across these DMAs, with a bit less than one-quarter of self-response predicted to come online in the Greenwood-Greenville (Miss.) DMA and two-thirds of self-response predicted to come online in Miami-Ft. Lauderdale. This type of variation will be important in the development of messaging and communications strategy to promote self-response from areas that are predicted to have lower rates of self-response.

Table 4. The 10 DMAs With the Lowest Predicted Self-Response to the 2020 Census

DMA	State(s)	2020-PSRS	2020-IPSR
Harlingen-Weslaco-Brownsville-McAllen	Texas	38%	48%
Laredo	Texas	38%	51%
Yuma-El Centro	Ariz., Calif.	45%	5%
Greenwood-Greenville	Miss.	45%	22%
Corpus Christi	Texas	47%	53%
El Paso (Las Cruces)	NM, Texas	48%	59%
Meridian	Ala., Miss.	49%	25%
Fresno-Visalia	Calif.	49%	50%
Myrtle Beach-Florence	SC	50%	42%
Miami-Ft. Lauderdale	Fla.	50%	66%

4. Historical Timing Benchmarks

4.1 Methodology

The first step in better understanding variation in response to the 2020 Census is identifying factors that might affect the timing of self-response. The CRAT team accomplished this by examining previous Census Bureau data collections at the tract level, across demographic groups, and by response-mode and contact strategy.

4.1.1 Tract-Level Benchmarks

To assess tract-level patterns in self-response,[13] the CRAT team examined daily summary statistics for the tracts. To account for differences in the size of tracts, the CRAT team weighted these summary statistics using 2010 Census estimates of the number of households present in each tract. To evaluate the central tendencies and variances of daily tract-level self-response, the CRAT team calculated weighted quartiles, means, and variances and plotted these results to examine how the distribution of daily tract-level self-response changed in the 71,889 tracts for which 2010 self-response data are available.

4.1.2 Demographic Benchmarks

To better understand how demographic groups might differ in the timing of their response to the 2020 Census, the CRAT team examined the timing of response to the 2010 Census for a number of demographic groups identified by communications professionals and summarized in Table 5. Because there was no public reporting of participation rates by demographic groups, the CRAT team produced daily cumulative response rates using the 2010 Census Edited File. The CRAT team produced these rates through the period of Nonresponse Followup. Including Nonresponse Followup operations in the demographic benchmarks gives a point of comparison for campaign optimization during the later stages of the 2020 Census. Because the demographic benchmarks include responses gathered by enumerators they are response rates rather than self-response rates.

The CRAT team examined daily response rates for these demographic groups at two geographic levels, including national and census division. The correlated nature of daily response rates made disclosure avoidance procedures, other than those appropriate to large geographic areas, infeasible.[14] This report will focus on national-level response rates among different demographic groups.

---

[13] The CRAT team produced tract-level self-response rates for current census tract geographies using publicly reported participation rates for tract geographies that were current at the time of the 2010 Census enumeration. For details on this process please see Technical Appendix B.3.

[14] Under a differential privacy framework, the size of the privacy risk and, therefore, the amount of noise injection required increases when estimates are related (Dwork and Roth, 2014). Cumulative daily rates from the same geography are necessarily related. In addition, the use of avoidance disclosure procedures appropriate to smaller geographic areas would have restricted us to demographic variables that produced mutually exclusive groupings. In such groupings, two different statistics could not include the same individual. The CRAT team followed the

Working with communications professionals (i.e., campaign optimization and media planning experts), the CRAT team identified the broad groupings of demographic characteristics in Table 5 as being both potentially valuable to the communications campaign and included on the 2010 Census questionnaire. Some of these characteristics, such as homeownership status, are properties of the entire household. Others, such as sex, age, or race, are individual characteristics. For individual characteristics, the CRAT team assigned responses to demographic groups using the characteristics of the individual listed as Person 1 on the 2010 Census questionnaire.

Table 5. Demographic Characteristics of Interest to the Communications Campaign

Demographic Characteristic
Sex
Age
Race
Homeowner/Renter
Marital Status
Whether or Not They Have Children
Hispanic Origin
Usual Residence

After assigning responses to different demographic groups, the CRAT team produced cumulative daily response rates for each group and indexed these rates to the final rate of response for each group. Indexing response rates to produce relative response rates removes differences in the final rate of response among these groups and makes it possible to compare the timing of response independent of the final level of response. Different demographic groups respond at different rates. By indexing to the final rate of response, the relative rate of response on a specific day of the enumeration makes it possible to compare the speed of response. For example, if, on the 15th day of the enumeration, one group has a higher proportion of its final response than another does, it has responded more quickly at this point. The CRAT team examined these relative response rates graphically by producing curves that show how relative response rates increased over time during the 2010 Census enumeration.

4.1.3 Test Benchmarks

Examining the timing of self-response for two large-scale census tests, the 2015 NCT[15] and the 2017 Census Test, gives insight into how contact strategy and response mode affect the timing

appropriate disclosure avoidance procedure for large geographic areas, rounded all rates to four digits, and suppressed any statistic based on three or fewer observations (i.e., the rule of three). The Disclosure Review Board approved the timing benchmarks on July 16, 2018, with approval number CBDRB-FY-382.

[15] The CRAT team considered six out of nine panels from the 2015 NCT that are slight variations on the Internet First contact strategy as receiving the Internet First contact strategy. Two panels that were not included had more

of self-response. The CRAT team calculated the date of earliest valid submission, the contact strategy received, and the mode of self-response for each response to these two tests. After grouping responses by contact strategy and response mode, the CRAT team produced daily cumulative self-response rates. To ensure that valid results were produced while adhering to necessary disclosure avoidance procedures, the CRAT team produced these curves at the national level.[16] Both the absolute and relative response rates of these groupings were calculated and examined graphically to understand how contact strategy and response mode changed the timing of self-response.

4.2 Results

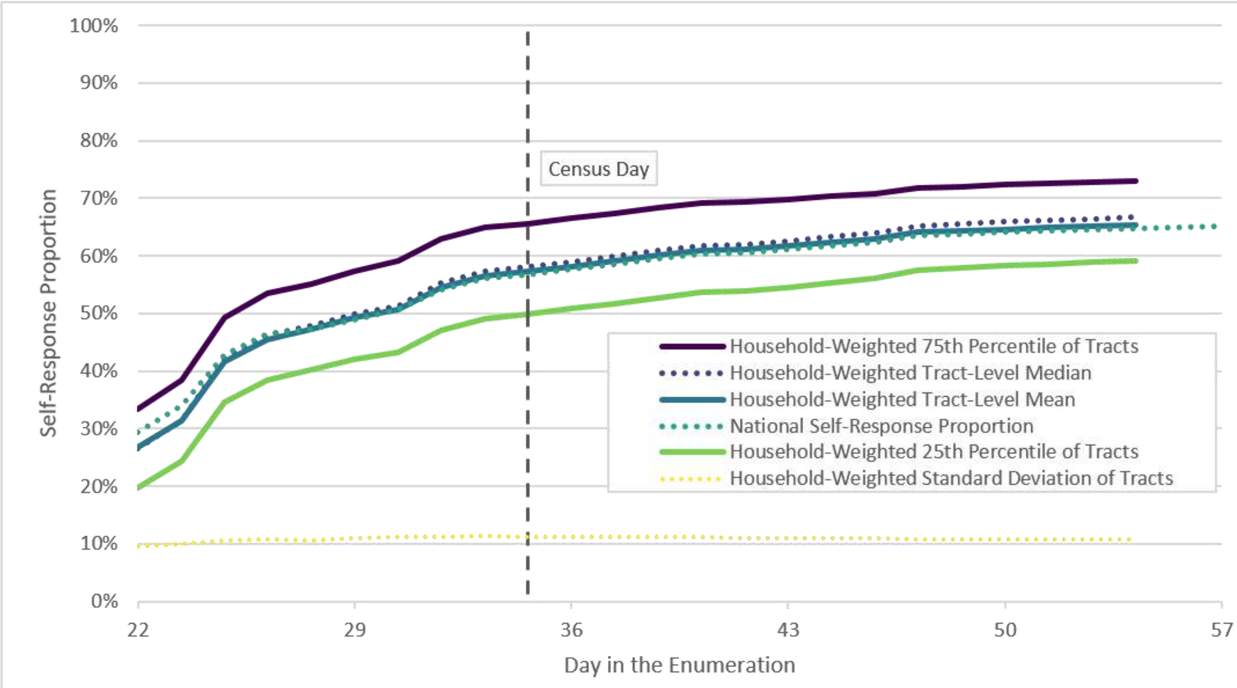
4.2.1 Tract-Level Benchmarks

For the period that this report examines, the summary statistics graphed in Figure 8 reveal a slowly declining increase in self-response. The mean tract-level response is slightly below the median, reflecting the existence of outlier tracts with relatively low self-response rates. The household-weighted interquartile range is fairly consistent across the entire period. The average of the daily interquartile range over the entire period is 15 percent. There is a slight increase in standard deviation leading to Census Day (April 1, 2010) and then declining afterwards. Examined at an aggregate level, these data do not show a strong temporal pattern in the variance of tract-level self-response.

---

significant deviations from the control Internet First contact strategy. Panel 4 received a paper questionnaire two weeks earlier than is typical of an Internet First contact strategy, and Panel 8 never received a paper questionnaire. When using self-response data from the 2015 NCT to develop timing benchmarks for the 2020 Census, the CRAT team excluded these two panels from consideration (Phelan, 2016).

[16] The CRAT team complied with disclosure avoidance procedures appropriate to national-level statistics by rounding rates to four digits and excluding all rates based on three or fewer observations. For more information on disclosure avoidance procedures, see Dwork and Roth (2014). The Disclosure Review Board approved the timing benchmarks on July 16, 2018, with approval number CBDRB-FY-382.



Source: 2010 Census Take 10 Tract-Level Participation Rates

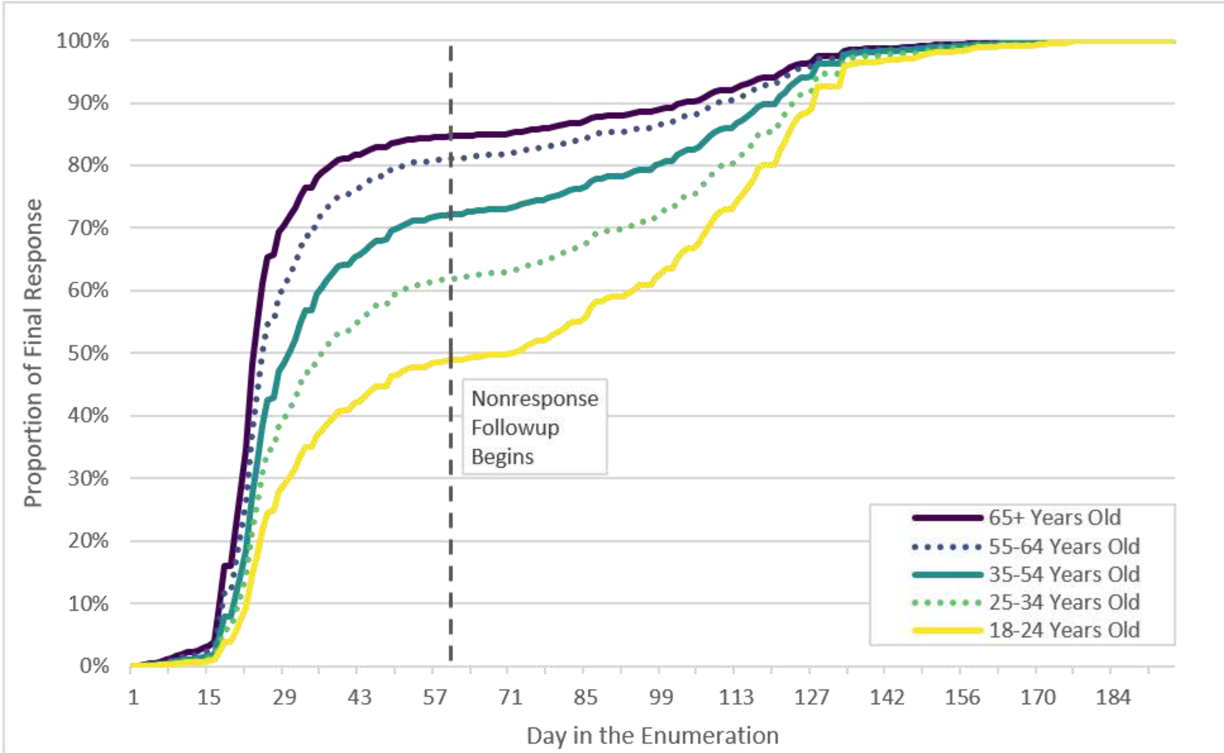
Figure 8. Cumulative Self-Response Rates to the 2010 Census

4.2.2 Demographic Benchmarks

The CRAT team collected response rates to the 2010 Census for more than 100 different demographic groups, but this report shows relative response curves for a smaller subset. Although the CRAT team did examine the response timing of groups defined using multiple demographic characteristics, this report presents the response pattern results only of groups defined by a single demographic characteristic in order to provide clear insights and to avoid presenting an overwhelming number of curves.

To ensure response timing is comparable across groups, relative response rates were used. A relative response rate of 25 percent indicates that one-quarter of the people in that group who will respond have done so by that day. Higher rates of relative response earlier in enumeration indicate that the group responds more quickly, while slower responding groups only show large increases in relative response rates towards the end of the enumeration.

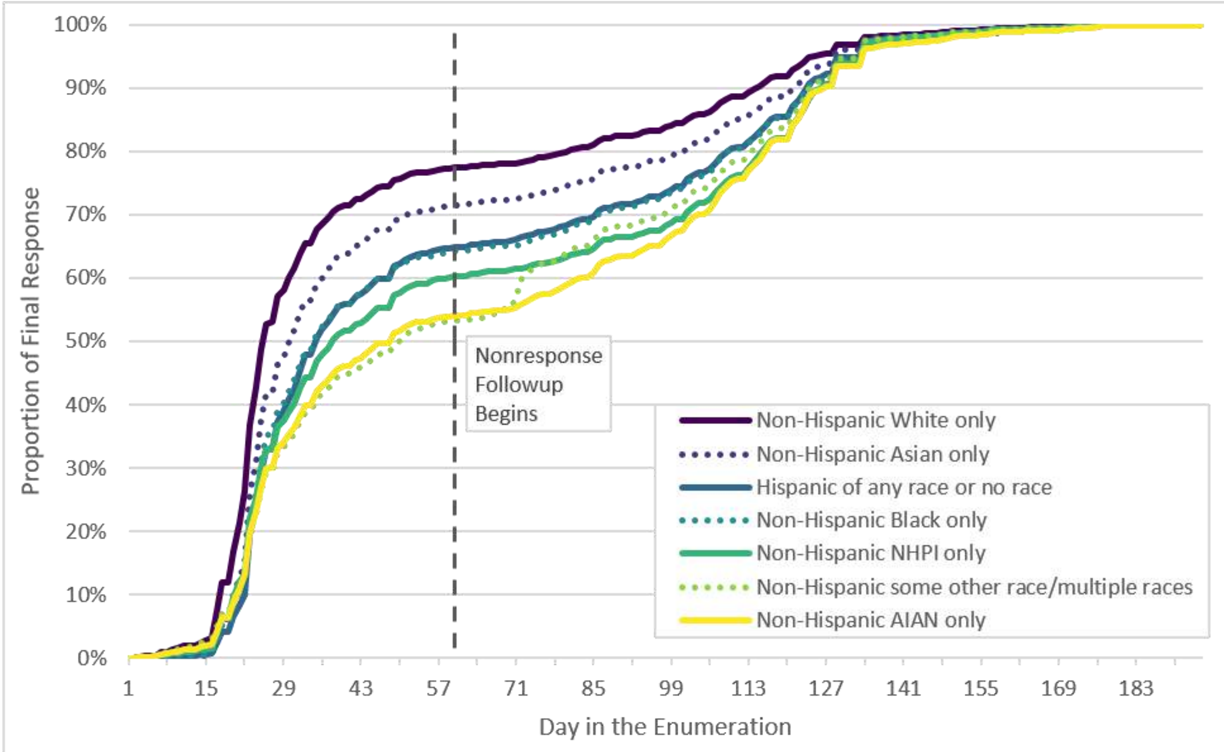
The most consistent pattern in relative response timing is that older people respond more quickly than younger people. When combining age group with each of the major race and Hispanic-origin groups, the same pattern is present, as it is when combining age with gender. Figure 9 shows response timing by age group and the basic shape of relative response rate curves. There is an initial acceleration in relative rates, which is higher for groups that respond more quickly. There is a later increase in response rates, which is slight for groups that respond quickly but more pronounced for groups that are slow to respond. Because these are relative response rates, all curves converge at 100 percent.



Source: 2010 Census Edited File

Figure 9. Relative Cumulative Response Rates to the 2010 Census by Age Group of Householder

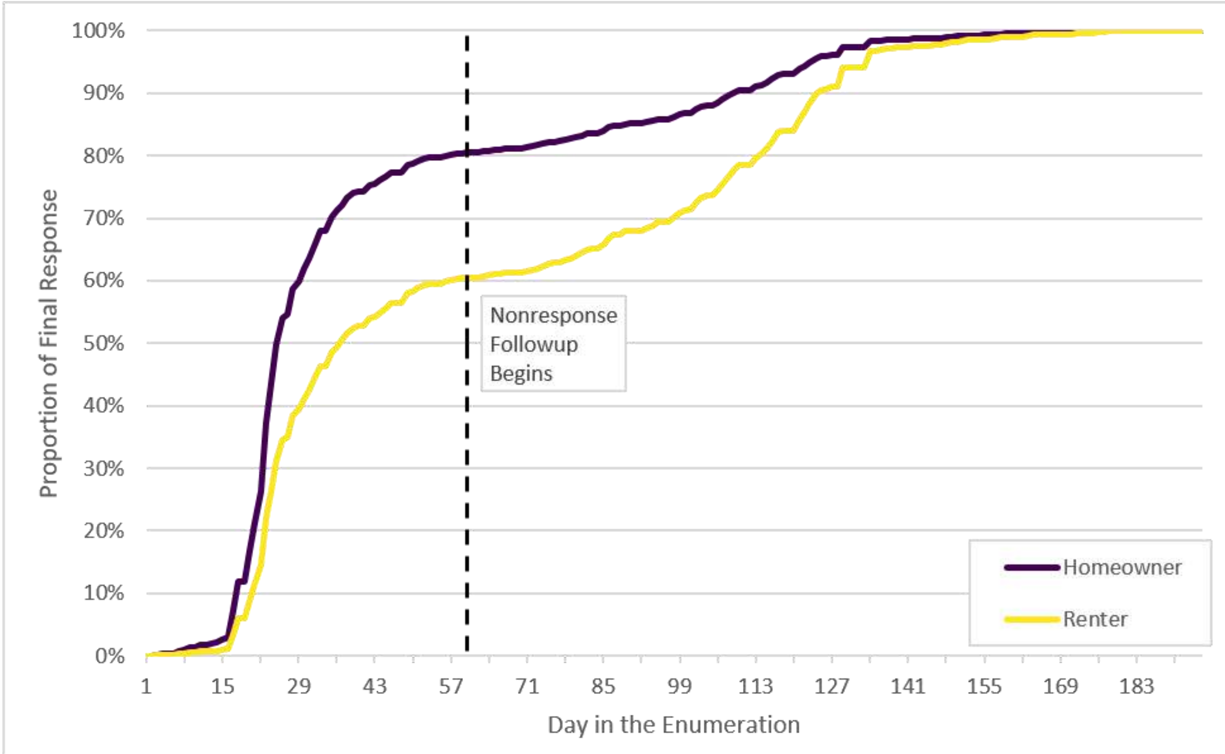
Another demographic characteristic that shows a clear pattern in response timing is race and Hispanic-origin groups. Looking at Figure 10, there is a clear distinction between different race and Hispanic-origin groupings. Non-Hispanic White householders respond most quickly, and non-Hispanic Asian householders are slightly slower to respond. The other race and Hispanic-origin groupings respond even slower with non-Hispanic American Indian and Alaska Native (AIAN) householders having the slowest relative response rate. None of the race and Hispanic-origin groups match the pace of respondents age 65 and over. As the curve in Figure 10 shows, even non-Hispanic White householders—which have the highest rates of relative response through the entire enumeration—have a more pronounced second acceleration in relative response rate compared with adults 65 and older as seen in Figure 9.



Source: 2010 Census Edited File

Figure 10. Relative Cumulative Response Rates to the 2010 Census by Race and Hispanic Origin of Householder

Another demographic characteristic that influenced relative response timing to the 2010 Census was homeownership. Homeowners are associated with a faster pattern of relative response (see Figure 11).



Source: 2010 Census Edited File

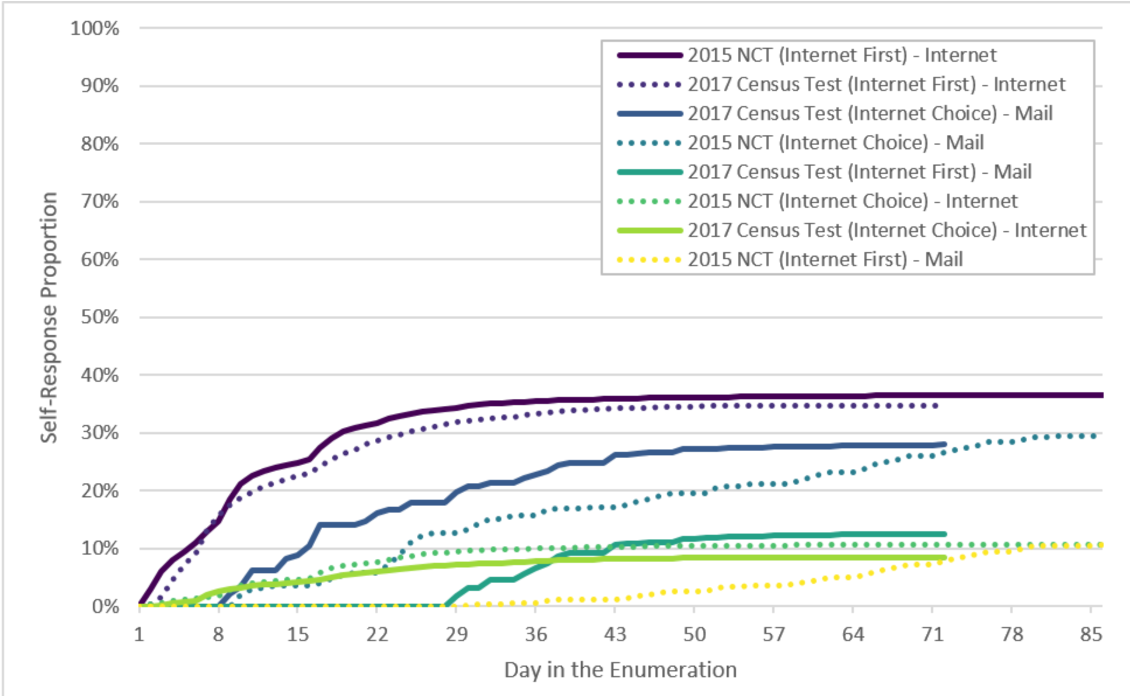
Figure 11. Relative Cumulative Response Rates to the 2010 Census by Homeownership

4.2.3 Test Benchmarks

The timing of self-response to the 2015 NCT and the 2017 Census Test show clear differences in the amount and speed of self-response by contact strategy and response mode. The daily absolute response rates shown in

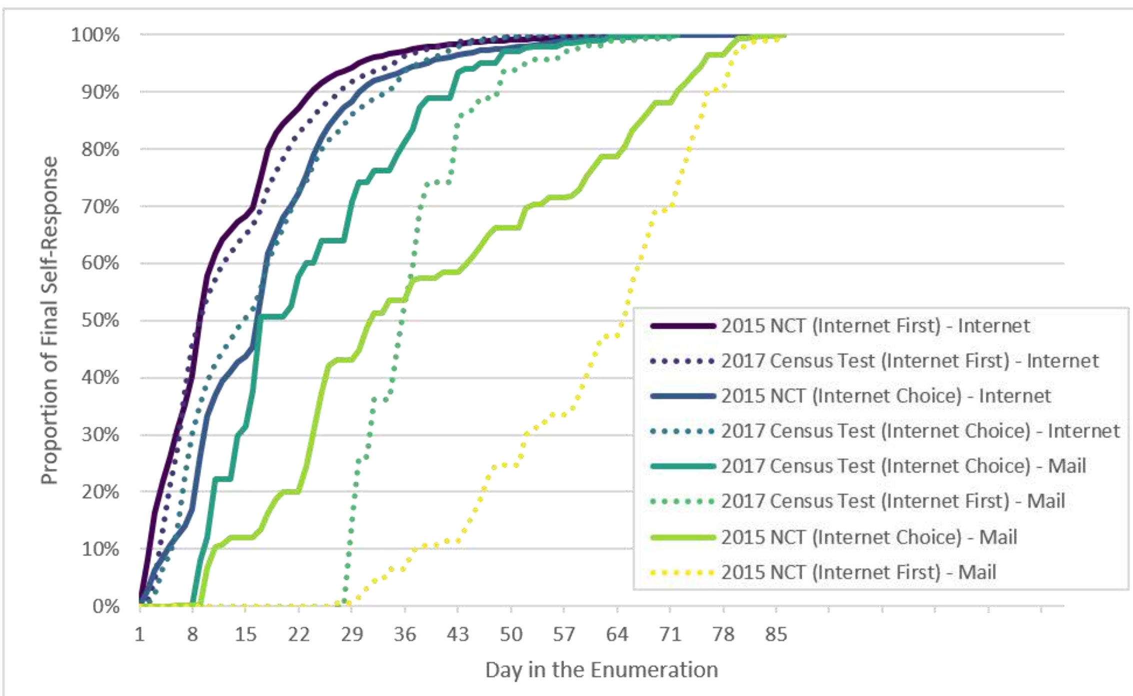
Figure 12 demonstrates the interaction between contact strategy and response mode. For both tests, the internet was the dominant response mode for the Internet First contact strategy, and mail predominated for the Internet Choice contact strategy. There is greater similarity in the pattern of self-response for the internet response modes than for the mail response mode. The 2015 NCT has a slower mail response rate than the 2017 Census Test.

Figure 13 shows the similarity of the timing of internet responses regardless of contact strategy when examining relative self-response rates. The final amount of self-response by internet under the Internet Choice contact strategy is lower for both tests, but a similar percentage of final self-response by internet comes in on each day regardless of contact strategy. The slower pace of mail response to the 2015 NCT is visible when examining the relative rates as well.



Source: 2015 National Content Test (Panels 1, 2, 3, 5, 6, and 9) and 2017 Census Test

Figure 12. Absolute Response Rates by Census Test, Contact Strategy, and Response Mode



Source: 2015 National Content Test (Panels 1, 2, 3, 5, 6, and 9) and 2017 Census Test

Figure 13. Relative Response Rates by Census Test, Contact Strategy, and Response Mode

5. Mindsets

5.1 Methodology

To help with creative message development, ad campaigns commonly develop psychographic profiles (or "mindsets") of their customers according to their knowledge, attitudes, and practices towards a particular product. For the 2020 ICC, the customers are every single person residing in the U.S. on April 1, 2020 and the product is the 2020 Census. Therefore, the CRAT team has produced six mindsets that reflect shared patterns of attitudes, behaviors, and motivators toward the 2020 Census.

5.1.1 Data Sources

The Census Bureau fielded the 2020 CBAMS Survey from February 20 through April 17, 2018. The 2020 CBAMS Survey included both an online and mail response mode and had a sample size of 17,283. The questionnaire consisted of 61 questions assessing a respondent's knowledge and attitudes as they pertain to participation in the 2020 Census and included nominal, ordinal, and interval answer formats. The 2020 CBAMS sampling strategy ensured a sufficient number of responses from various subpopulations. The 2020 Census Barriers, Attitudes, and Motivators Study Survey Report contains a more complete description of the 2020 CBAMS Survey methodology and questionnaire (McGeeney et al., 2019).

The Census Bureau administered the 2020 CBAMS survey, so Title 13 of the U.S. Code protects the original results. The CRAT team produced CBAMS mindsets using these results and appended them prior to the creation of the Public Use Microdata Sample (PUMS). Disclosure avoidance procedures applied to create the PUMS version of the 2020 CBAMS results ensure that the Census Bureau can publicly share all data products without risk of disclosing Title 13 data.[17] The descriptions of mindsets included in this report were produced using the PUMS data set. For more details on the PUMS data set, see the 2020 Census Barriers, Attitudes, and Motivators Study Survey Report (McGeeney et al., 2019).

5.1.2 Data Preparations

The CRAT team transformed the 2020 CBAMS Survey results into variables for use in clustering.[18] Since the number of potential inputs to mindset segmentation was large, the CRAT team used principal components analysis (PCA) to identify components that capture most of the variance in 2020 CBAMS Survey responses. PCA also ensured that variables used for mindset creation were not correlated with each other, which could increase the relative weight of a single, underlying characteristic by counting it multiple times.

Applying PCA involved decomposing the correlation matrix into a set of orthogonal linear combinations. In the PCA results, the first component accounts for the highest proportion of variability in the data, and each successive component accounts for the highest proportion of

---

[17] The Disclosure Review Board approved the Public Use Microdata Sample version of the original 2020 CBAMS Survey for release on August 13, 2018 with approval number CBDRB-FY18-422.

[18] The CRAT team used predictive mean matching hot deck imputation to ensure complete cases in the CBAMS data file (Andridge & Little, 2010).

remaining variability. The fraction of variance represented by each component is proportional to each component's eigenvalue. In our case, the optimal number of principal components was eight, as indicated by a scree plot. The added information from each additional component diminished after that point. The CRAT team then used varimax rotation, which maximizes the sum of the variances of the squared loadings to ensure that each variable corresponds to a single factor, so that components were directly interpretable. Finally, mean factor scores across the eight factors were calculated for each case. Technical Appendix B.4 contains the list of variables input to the PCA as well as the factor loadings.

5.1.3 Creating Candidate Mindset Solutions

The mean factor scores derived from the PCA served as inputs into the clustering algorithms. The CRAT team used two clustering algorithms, k-means and Ward's method, to produce three candidate mindset solutions, from which the final was selected. Technical details of this process are contained in Technical Appendix B.6.

5.1.4 Selecting the Final Mindset Solution

The CRAT team presented three of the mindset solutions that it produced using k-means and Ward's method to ICC stakeholders and, from those, selected the final mindset solution. The three candidate solutions considered were:

- Solution A, which used k-means to produce five mindsets
- Solution B, which used k-means to produce six mindsets
- Solution C, which used Ward's method to produce six mindsets

All three candidate solutions had two similarities. Across the three solutions, there were consistent top "motivators," or reasons to participate in the census, including "It helps determine funding for public services in my community like schools and fire departments," and "It is my civic duty (along with voting, jury duty, paying taxes)."[19] In all three candidate mindset solutions, there was one mindset that tended to skew older, with a significantly larger percentage of respondents 65 years or older.

However, the candidate mindset solutions also had important differences. Although candidate Solution A offered the simplest solution, with only five mindsets, most stakeholders believed that this mindset solution did not differentiate enough across racial and Hispanic-origin groups. Solutions B and C both offered a more nuanced set of six mindsets. Although the mindset groups in candidate Solution B were the most evenly sized by the number of respondents in each solution, the mindsets in Solution C differentiated the most on key outcomes of interest, including demographic characteristics and attitudes toward and knowledge about the census.

---

[19] Based on the 2020 CBAMS Survey question, "Which ONE of the following is the most important reason, to you personally, that you should fill out the census form? (a) It helps determine funding for public services in my community like schools and fire departments, (b) It determines how many elected representatives my state has in Congress, (c) It is used to enforce civil rights laws, (d) It provides information for my local government to plan for changes in my community, (e) It shows that I am proud of my cultural heritage, (f) It is my civic duty (along with voting, jury duty, paying taxes), (g) It contributes to a better future for my community."

When comparing Solutions B and C, all stakeholders thought that Solution C was a more accurate representation of the population as they understand it. Team Y&R's multicultural partners felt that Solution C best reflected their unique audience's demographics, attitudes, knowledge, and interests. Similarly, the digital advertising partners cited the granularity of differentiation reflected in Solution C as important for mapping mindsets to media data and for creating targeted messaging and media products for the campaign.

5.2 Results

5.2.1 Introduction to Mindsets

After the Census Bureau and communications professionals selected Solution C as the final mindset solution, the CRAT team appended them to the CBAMS responses, produced the PUMS data set, and developed in-depth profiles using the PUMS for each of the six mindsets structured around the following four questions:

- **Who are they?** The CRAT team focused on the demographic characteristics of each mindset, including age, race and Hispanic origin, income, homeownership, presence of children in the household, marital status, internet use, English proficiency, and country of birth.
- **Do they intend to respond, and how do they think about the census?** The CRAT team examined what percentage of each mindset said they intend to respond to the census, their levels of knowledge about the census (high: 8 to 11 knowledge questions answered correctly, medium: 5 to 7 correct, or low: 0 to 4 correct), and their levels of civic participation (high: 7 to 10 activities, medium: 4 to 6 activities, or low: 0 to 3 activities).
- **What are their potential barriers to participation?** The CRAT team examined the percentage of each mindset that exhibited attitudes or behaviors that could dampen self-response to the census. These attitudes and behaviors included concern about confidentiality, fear of repercussions, distrust in the government, perceptions of whether or not it matters if they are counted in the census, and perceptions of whether or not participating will benefit or harm their communities or them personally.
- **What are their potential motivators for participation?** The CRAT team examined which motivators, or reasons to participate in the census, resonated with the largest percentage of respondents in each mindset. These motivators were analyzed for each mindset in two ways: (1) the single most important reason to participate in the census among a list of options (see Table 6) and (2) the importance of each motivator (see Table 7).

Table 6. Single Most Important Motivator (Forced Choice)

Reasons to Participate in the 2020 Census	2020 CBAMS Survey Average
It helps determine funding for public services in my community like schools and fire departments	30%
It is my civic duty (along with voting, jury duty, paying taxes)	25%
It contributes to a better future for my community	17%

It provides information for my local government to plan for changes in my community	15%
It determines how many elected representatives my state has in Congress	9%
It is used to enforce civil rights laws	2%
It shows that I am proud of my cultural heritage	1%

Source: 2020 CBAMS Public Use Microdata Sample

Table 7. Importance of Individual Motivators

Reasons to Participate in the 2020 Census	**2020 CBAMS Survey Average (% Extremely/Very important)**
Hospitals and healthcare	94%
Fire departments	94%
Roads and highways	92%
Police departments	92%
That civil rights laws are enforced	86%
Schools and the education system	85%
Fulfilling your civic duty	82%
Contributing to a better future for your community	81%
Providing information for your local government to plan for changes in your community	76%
Determining how many elected representatives your state has in Congress	72%
Job training programs	61%
Public transportation	57%
Showing you are proud of your cultural heritage	56%
Daycare for children	50%

Notes: 2020 CBAMS Survey questions about the importance of each motivator follow this pattern: "How important, if at all, is each of the following to you personally? (a) Extremely important, (b) Very important, (c) Somewhat important, (d) Not too important, (e) Not at all important." The percentages reported for each motivator reflect answers of "Extremely important" or "Very important."

Figure 14 is a visual dashboard of key characteristics of the final six mindsets, which, in order of the percentage who intend to respond, are: (1) Eager Engagers, (2) Fence Sitters, (3) Confidentiality Minded, (4) Head Nodders, (5) Wary Skeptics, and (6) Disconnected Doubters. It is important to note that actual self-response rates tend to be lower than intended response rates. Research suggests that even individuals who report a commitment to participate may not follow through on their intention. Ajzen (1991) argues that there is often a gap between a person's intended and actual behavior, and unforeseen costs and circumstances ultimately prevent some people from carrying out their intended behavior.

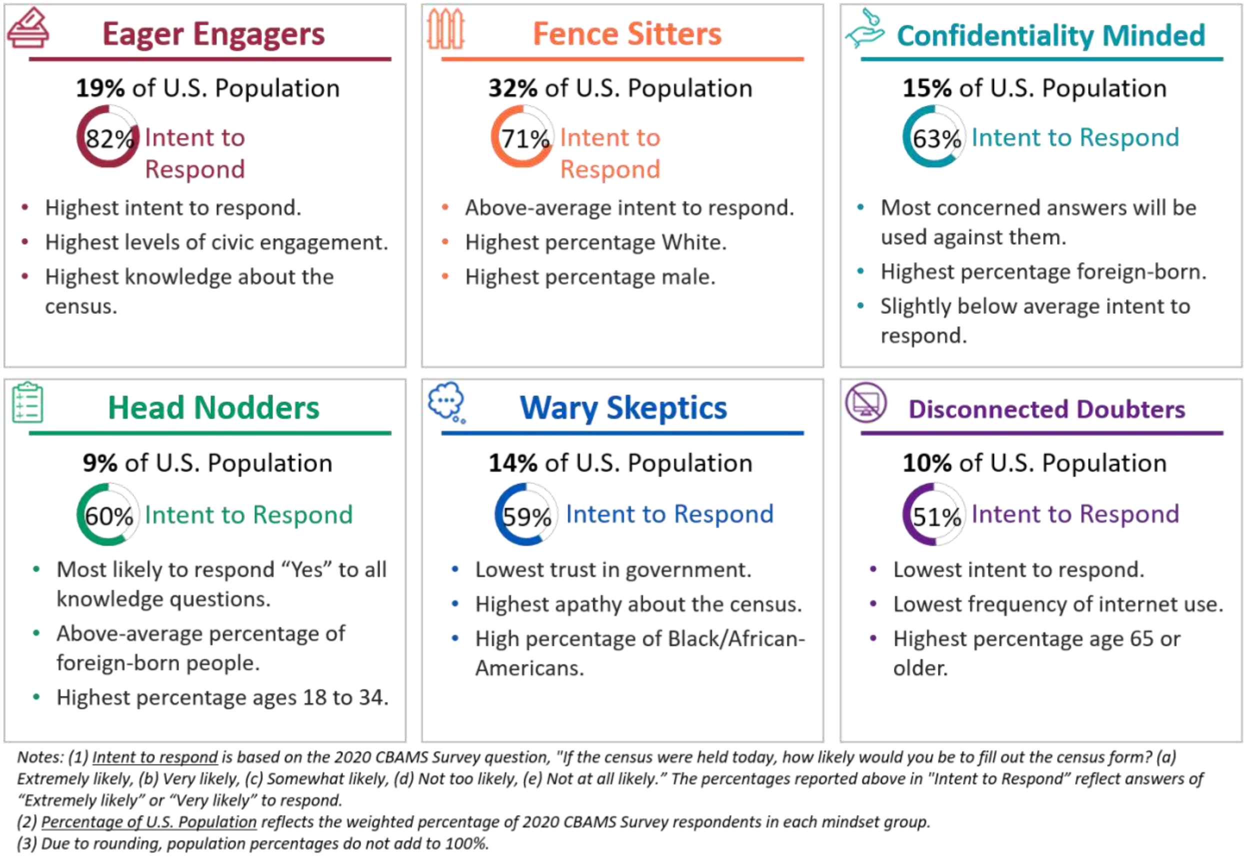


Figure 14. Overview of Mindsets

5.2.2 Detailed Mindset Profiles

5.2.2.1 Eager Engagers

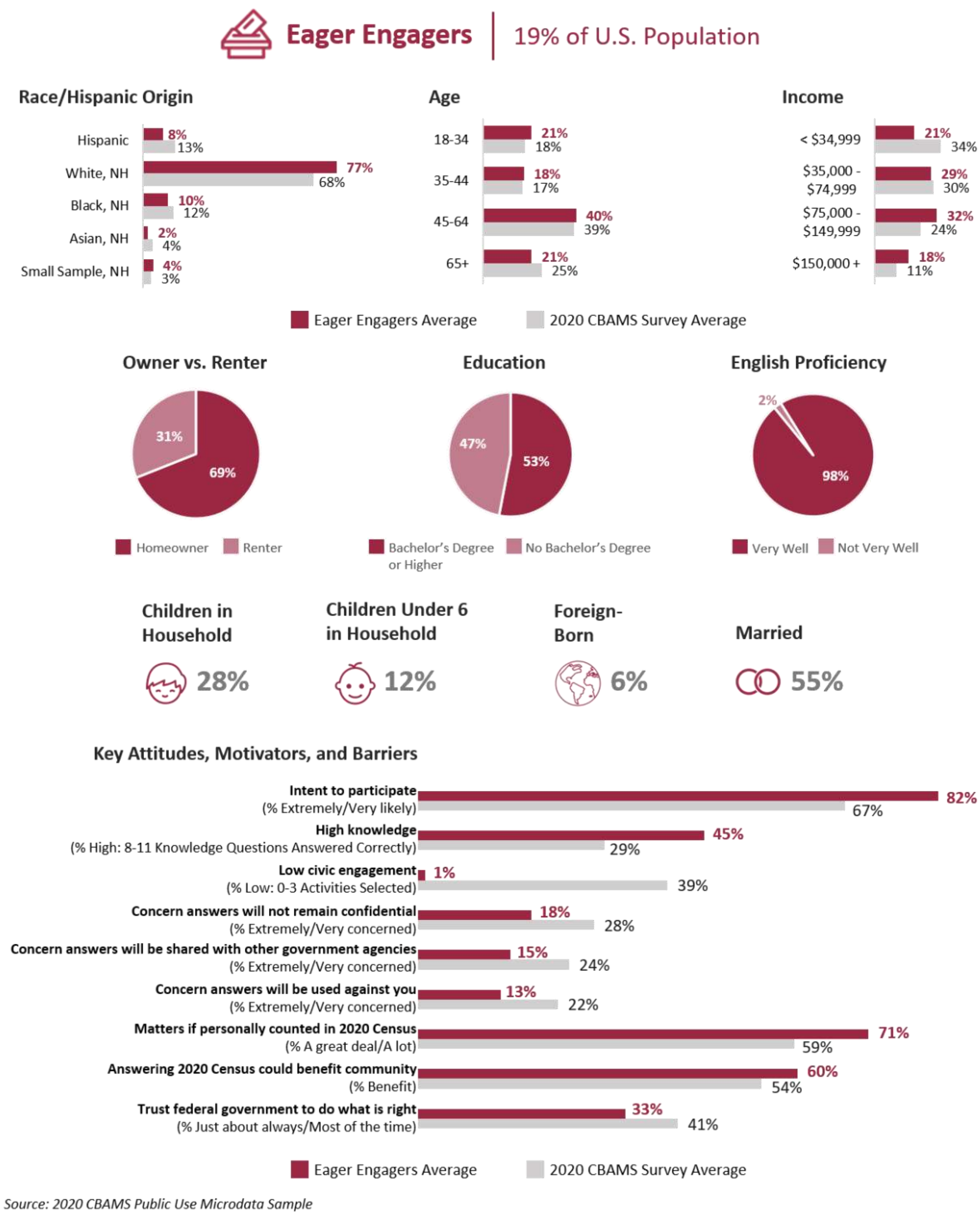


Figure 15. Eager Engagers Key Characteristics

Who are they?

As Figure 15 reports, 19 percent of the U.S. population is classified as Eager Engagers. Figure 15 also shows that the largest racial group in the Eager Engagers is non-Hispanic White (77 percent), followed by non-Hispanic Black or African American (10 percent), Hispanic (8 percent), non-Hispanic small-sample race (4 percent), and non-Hispanic Asian (2 percent).

In terms of age, the largest share of Eager Engagers are 45 to 64 (40 percent). However, younger (18 to 34, 21 percent) and older (65 or older, 21 percent) people are represented at equal rates. People 35 to 44 make up the smallest share of the Eager Engagers (18 percent). This mindset group also has the largest share of frequent internet users of any mindset, with 86 percent using the internet "almost constantly" or "several times a day."

Eager Engagers have the highest average household income of any mindset group, with the largest share of people (32 percent) living in households that earn $75,000 to $149,999 annually. Sixty-nine percent of Eager Engagers own their homes, while 31 percent are renters. As Figure 15 shows, there are slightly more married (55 percent) than nonmarried householders, and roughly 28 percent of Eager Engagers live in a household with children younger than 18. Finally, most Eager Engagers were born in the United States (94 percent) and speak English at a self-reported level of "very well" (98 percent).[20]

Do they intend to respond, and how do they think about the census?

As their name suggests, the Eager Engagers mindset has the highest percentage of people who said they intend to respond to the census (82 percent). As Figure 15 shows, those who have the Eager Engagers mindset also tend to display two characteristics related to high levels of intent to respond—knowledge about the census and high levels of civic participation. Nearly half (45 percent) of Eager Engagers are classified as having high knowledge and only 19 percent are classified as having low knowledge.

The majority of people in the Eager Engagers mindset have high levels of civic participation (73 percent), and nearly all (98 percent) have voted in an election. Ninety-one percent have signed a petition, and 87 percent have volunteered at an organization. A mere 1 percent of the Eager Engagers mindset was categorized as demonstrating low civic engagement.

What are their potential barriers to participation?

As Figure 15 demonstrates, the Eager Engagers mindset has a lower than average percentage of people who demonstrate beliefs or attitudes associated with low response. For instance, only 18 percent of Eager Engagers are concerned that the Census Bureau will not keep their answers to 2020 Census questions confidential,[21] and 15 percent are concerned that the Census Bureau

---

[20] English proficiency is based on the question, "How well do you speak English? (a) Very well, (b) Well, (c) Not well, (d) Not at all." Respondents are considered to have high English proficiency if they speak English at a self-reported level of "very well."

[21] Based on the question, "How concerned are you, if at all, that the Census Bureau will not keep answers to the 2020 Census confidential? (a) Extremely concerned, (b) Very concerned, (c) Somewhat concerned, (d) Not too

will share their answers with other government agencies.[22] Similarly, only 13 percent are concerned their answers to 2020 Census questions could be used against them.[23]

When asked whether they believe it matters if they personally are counted in the 2020 Census, 71 percent of Eager Engagers said they believed it mattered “a great deal” or “a lot”—the highest of any mindset—and 60 percent of those in the Eager Engagers mindset said they believed completing the 2020 Census form could benefit their community.

Although the Eager Engagers mindset did not have the highest levels of trust in government among the mindset groups, over half (52 percent) said they could trust their local government to do what is right “just about always” or “most of the time.” Forty percent and 33 percent said they could trust their state and the federal governments, respectively.

What are their potential motivators for participation?

When asked to identify the single most important reason for them, personally, to complete the census form, the largest share of Eager Engagers identified “It helps determine funding for public services in my community like schools and fire departments” followed by “It is my civic duty (along with voting, jury duty, and paying taxes).”[24]

When asked to rate the importance of each motivator, the Eager Engagers rated “determining how many elected representatives your state has in Congress” as “very important” or “extremely important” at a higher percentage than any other mindset group. As Figure 16 demonstrates, over 90 percent of Eager Engagers believed that eight of the 14 potential motivators were either “very important” or “extremely important.”[25] However, compared to

---

concerned, (e) Not at all concerned." Respondents are considered to be concerned about confidentiality if they responded “extremely concerned” or “very concerned.”

[22] Based on the question, “How concerned are you, if at all, that the Census Bureau will share answers to the 2020 Census with other government agencies? (a) Extremely concerned, (b) Very concerned, (c) Somewhat concerned, (d) Not too concerned, (e) Not at all concerned." Respondents are considered to be concerned that their data will be shared with other government agencies if they responded “extremely concerned” or “very concerned.”

[23] Based on the question, “How concerned are you, if at all, that the answers you provide to the 2020 Census will be used against you? (a) Extremely concerned, (b) Very concerned, (c) Somewhat concerned, (d) Not too concerned, (e) Not at all concerned." Respondents were considered to be concerned that their answers will be used against them if they responded “extremely concerned” or “very concerned.”

[24] Based on the question, “Which ONE of the following is the most important reason, to you personally, that you should fill out the census form? Select only one answer. (a) It helps determine funding for public services in my community like schools and fire departments, (b) It determines how many elected representatives my state has in Congress, (c) It is used to enforce civil rights laws, (d) It provides information for my local government to plan for changes in my community, (e) It shows that I am proud of my cultural heritage, (f) It is my civic duty (along with voting, jury duty, paying taxes), and (g) It contributes to a better future for my community.”

[25] Based on the question “How important, if at all, is each of the following to you personally? (a) Extremely important, (b) Very important, (c) Somewhat important, (d) Not too important, (e) Not at all important.” Respondents were asked to consider the following: (1) Daycare for children, (2) Fire departments, (3) Police departments, (4) Hospitals and healthcare, (5) Job training programs, (6) Roads and highways, (7) Public transportation, (8) Schools and the education system, (9) Showing you are proud of your cultural heritage, (10) Contributing to a better future for your community, (11) Fulfilling your civic duty (for example, voting, jury duty, paying taxes), (12) That civil rights laws are enforced, (13) Determining how many elected representatives your

other motivators, a lower percentage of the Eager Engagers mindset group is motivated by daycare for children and pride in cultural heritage.

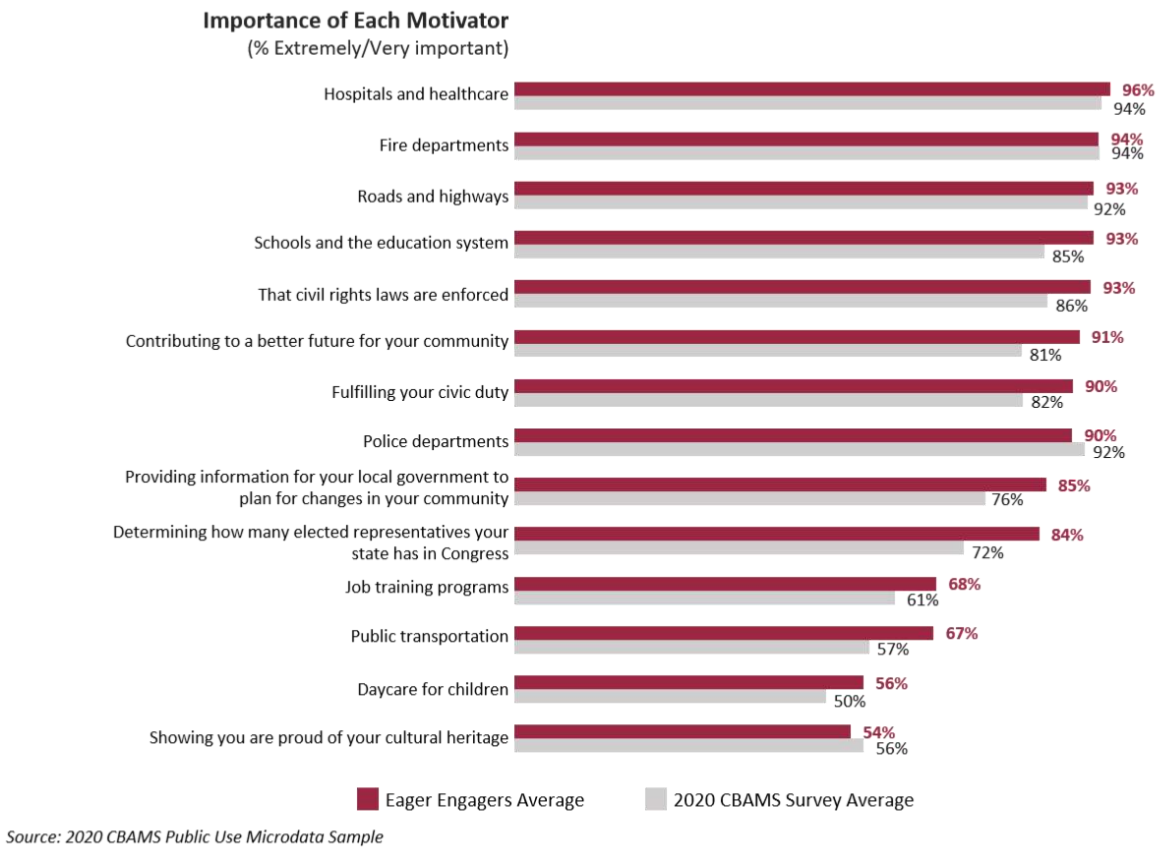


Figure 16. Eager Engagers Motivators

state has in Congress, and (14) Providing information for your local government to plan for changes in your community. Percentages reported here reflect the percentage of respondents who said that the item was “extremely important” or “very important” to them.

5.2.2.2 Fence Sitters

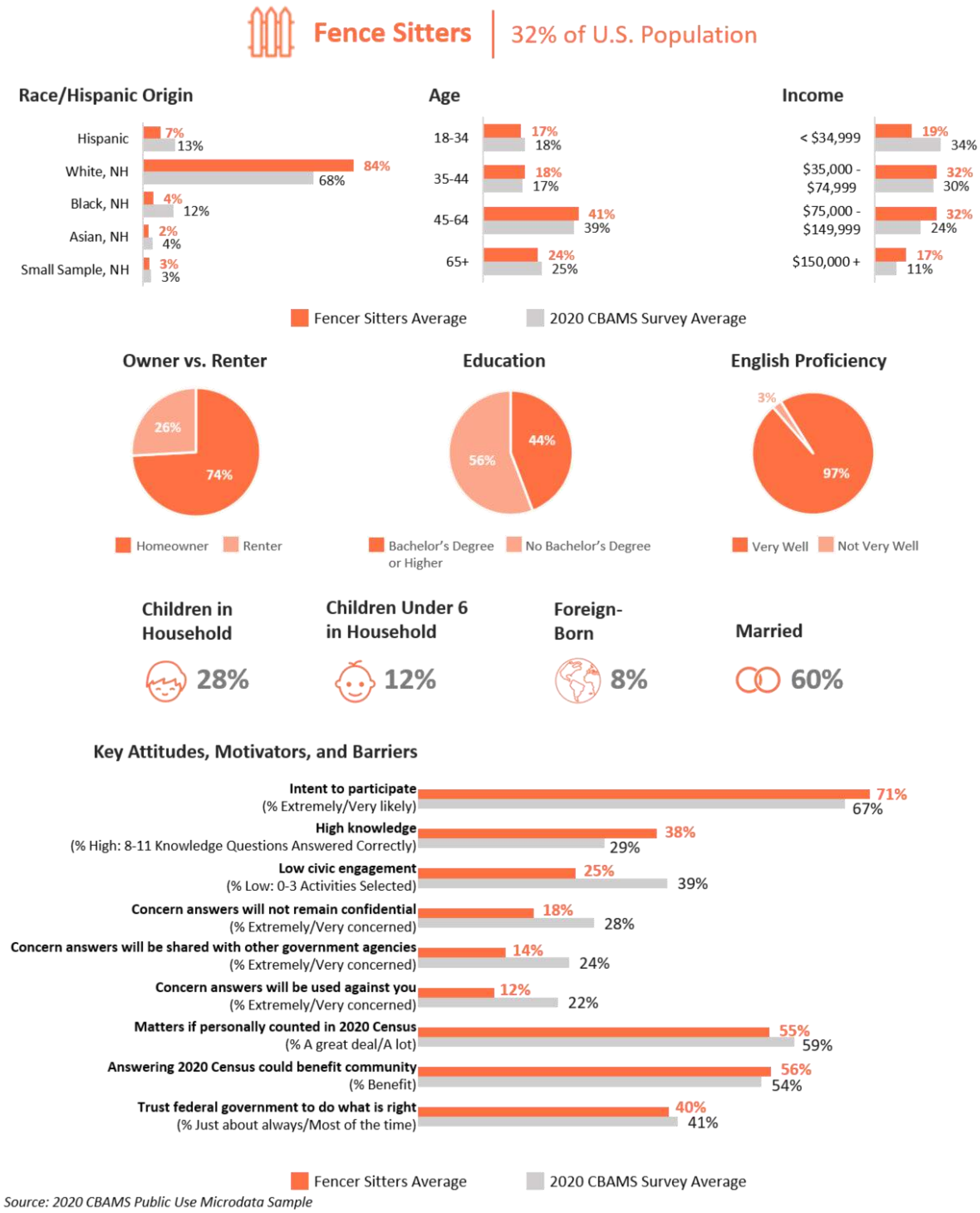


Figure 17. Fence Sitters Key Characteristics

Who are they?

Thirty-two percent of the U.S. population is classified as Fence Sitters. As shown in Figure 17, Fence Sitters are the least diverse mindset in terms of race and Hispanic-origin groups, with non-Hispanic White respondents making up 84 percent of the group. No other race and Hispanic-origin group makes up more than 7 percent of the Fence Sitters mindset: Hispanic origin (7 percent), non-Hispanic Black or African American (4 percent), non-Hispanic small-sample race (3 percent), and non-Hispanic Asian (2 percent).

Those of the Fence Sitters mindset also skew slightly older. People 45 to 64 make up 41 percent of the mindset group, followed by people 65 or older (24 percent), 35 to 44 (18 percent), and 18 to 34 (17 percent).

As Figure 17 demonstrates, the bulk of Fence Sitters live in households that earn between $35,000 and $149,000 annually (64 percent). Seventy-four percent own their homes, and 26 percent are renters. Like the Eager Engagers, 28 percent of Fence Sitters have children younger than 18 in the house, and over half (60 percent) of this mindset group are married. Members of this group also have an above-average frequency of internet use, with 80 percent using the internet "almost constantly" or "several times a day."

Almost all Fence Sitters are proficient in English (97 percent) and born in the United States (92 percent).

Do they intend to respond, and how do they think about the census?

Figure 17 also shows that an above-average percentage of Fence Sitters intend to respond to the 2020 Census (71 percent). This is the second-most knowledgeable mindset after the Eager Engagers group, with 38 percent of respondents classified as high knowledge. However, a slightly larger percentage of Fence Sitters (41 percent) are classified as medium knowledge, and 22 percent are classified as low knowledge. Nearly all (90 percent) know that the census is used to identify changes in the size, location, and characteristics of the people in the United States.

Roughly half (49 percent) of the Fence Sitters mindset were assigned to the medium civic engagement category. Ninety-two percent of Fence Sitters have voted in an election, 66 percent have signed a petition, and 66 percent have volunteered in an organization.

What are their potential barriers to participation?

People in the Fence Sitters mindset are not especially concerned that they will suffer negative consequences from participating in the census. Figure 17 shows that only 18 percent are "extremely concerned" or "very concerned" that their answers to the 2020 Census will not be kept confidential. Similarly, only 14 percent are concerned that their answers will be shared with other government agencies, and even less, 12 percent, are concerned that their answers will be used against them.

However, only 55 percent of the Fence Sitters group think it matters if they personally are counted in the 2020 Census. This is slightly below the average of 59 percent. Fifty-six percent believe the census will benefit their community, but when asked whether participating in the census would benefit or harm them *personally*, people were more inclined (52 percent) to say participating in the census would neither benefit nor harm them than that it would benefit them (33 percent).

Finally, Fence Sitters trust the federal government to do what is right “just about always” or “most of the time” at rates similar to the national average. Forty percent trust the federal government, 46 percent trust their state governments, and 56 percent trust their local governments.

What are their potential motivators for participation?

The largest share of Fence Sitters tended to cite “civic duty” (30 percent) or “helps determine funding for public services in my community” (28 percent) as the *one* most important reason to fill out the census form. When asked about the individual importance of each motivator, over 90 percent of Fence Sitters said that “fire departments,” “hospitals and healthcare,” “police departments,” and “roads and highways” were “extremely important” or “very important.” Figure 18 reports the percentage of Fence Sitters who said each motivator included in the 2020 CBAMS Survey was “very important” or “extremely important” to them.

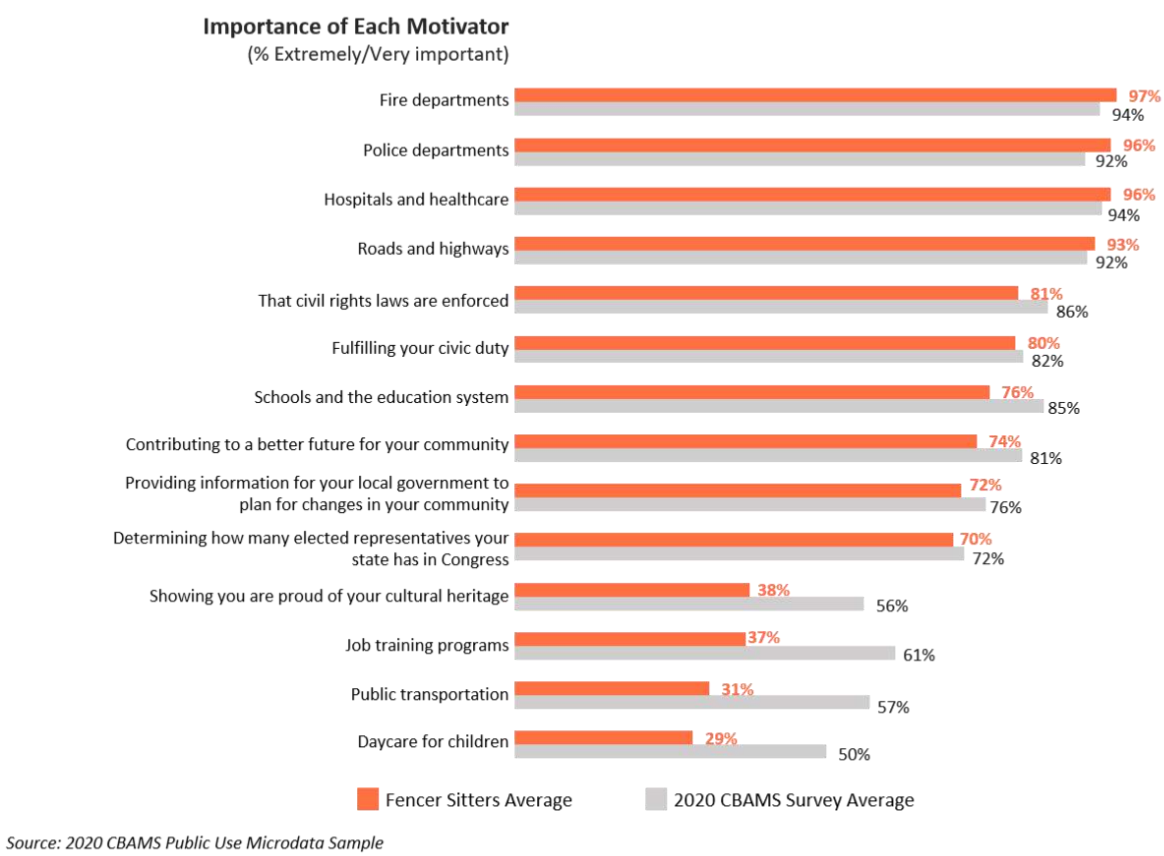


Figure 18. Fence Sitters Motivators

5.2.2.3 Confidentiality Minded

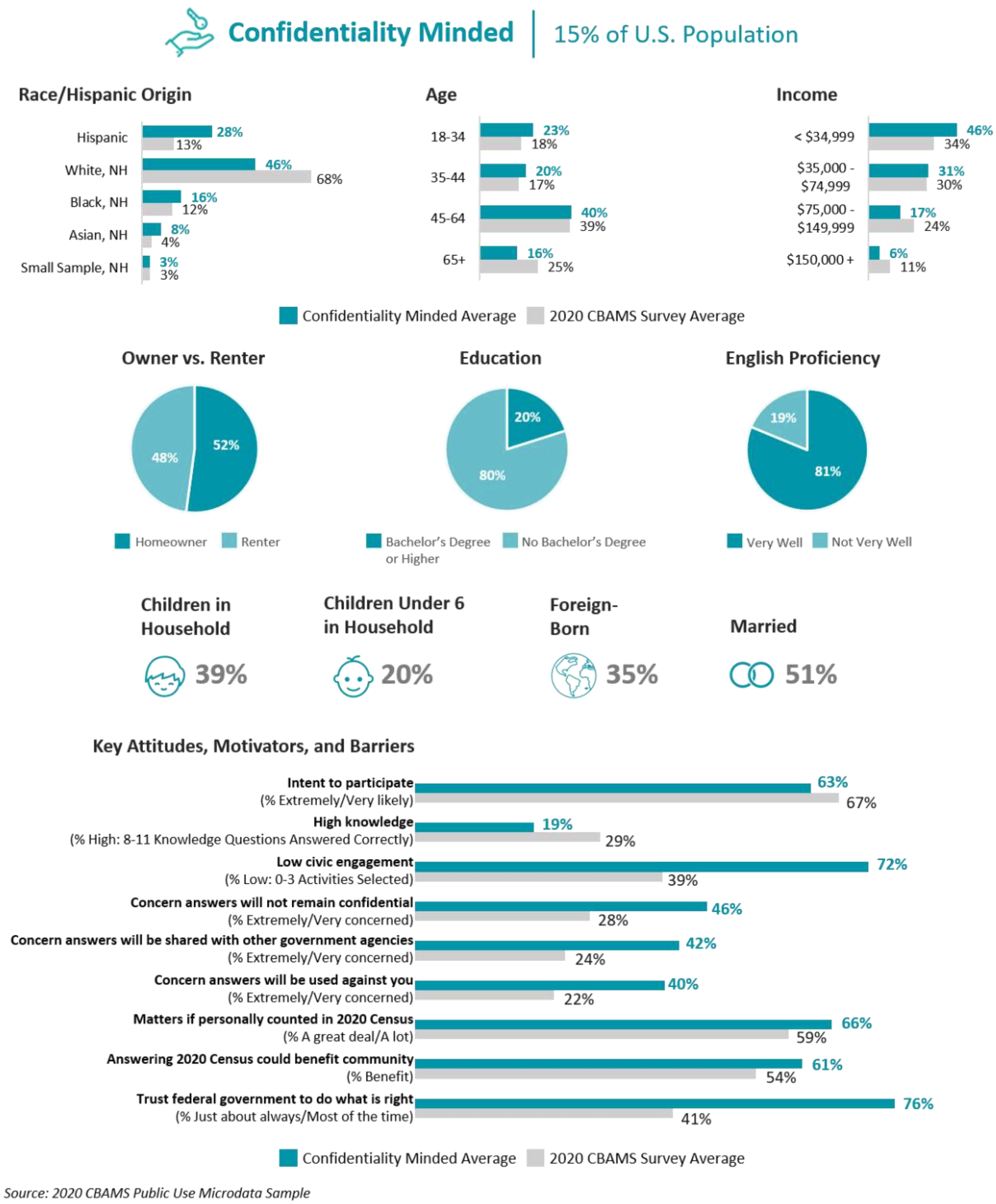


Figure 19. Confidentiality Minded Key Characteristics

Who are they?

Figure 19 shows that 15 percent of the U.S. population is in the Confidentiality Minded mindset group. The Confidentiality Minded mindset group is the most diverse mindset group with the highest percentage of foreign-born householders (35 percent). Figure 19 also shows that 46 percent of the Confidentiality Minded are non-Hispanic White, 28 percent are Hispanic, 16 percent are non-Hispanic Black or African American, 8 percent are non-Hispanic Asian, and 3 percent are non-Hispanic small-sample race. Nineteen percent of the Confidentiality Minded group do not speak English very well, making it the mindset group with the largest percentage of people not proficient in English.

The Confidentiality Minded mindset tends to skew slightly younger than the national average. Twenty-three percent of respondents in the Confidentiality Minded mindset group are 18 to 34 compared with the national average of 18 percent, and 20 percent are 35 to 44 compared with the national average of 17 percent. Following a similar pattern, 16 percent are 65 or older compared with the national average of 25 percent.

This mindset group has an average percentage of frequent internet users (67 percent).

As Figure 19 shows, the Confidentiality Minded are fairly evenly split between homeowners (52 percent) and renters (48 percent), and average income skews slightly lower than in other mindsets. Forty-six percent live in households that earn less than $34,999 annually, and 31 percent earn between $35,000 and $74,999 annually. Only 17 percent earn between $75,000 and $149,999, and very few (6 percent) live in households that earn more than $150,000 annually. More than one-third (39 percent) of respondents in the Confidentiality Minded mindset live in households with children 17 or younger, and roughly half (51 percent) are married.

Do they intend to respond, and how do they think about the census?

The Confidentiality Minded mindset group has a slightly below-average percentage of people who intend to respond to the census (63 percent), as shown in Figure 19. The largest share of Confidentiality Minded people (47 percent) are classified as low knowledge. This is the mindset group with the lowest percentage of respondents who know the census is used to determine the number of congressional representatives each state will have in Congress (31 percent), and it tied with the Disconnected Doubters for the group with the lowest percentage of people (26 percent) who know the census is used to decide how much money communities get from the government.

The Confidentiality Minded group has very low levels of civic engagement with 72 percent of respondents in this mindset classified as having low civic engagement. Only 60 percent have ever voted in an election.

What are their potential barriers to participation?

As the mindset name suggests, members of the Confidentiality Minded mindset tend to be significantly more concerned than other mindsets that their answers will not remain confidential, will be shared with other government agencies outside the Census Bureau, or will be used against them. Figure 19 shows that 46 percent of the Confidentiality Minded are "very concerned" or "extremely concerned" that the Census Bureau will not keep their answers confidential. Similarly, 42 percent are concerned that the Census Bureau will share their responses to the 2020 Census with other government agencies, and 40 percent are concerned their answers will be used against them.

Sixty-six percent of respondents in the Confidentiality Minded mindset group said it matters "a great deal" or "a lot" if they are personally counted in the 2020 Census. It is the mindset with the second-highest percentage of people who feel that it matters if they are counted. Similarly, 61 percent of the Confidentiality Minded believe their community could receive benefits for participating in the census—making them tied with the Eager Engagers as the mindset with the largest share of people who believe the census could benefit their community. The Confidentiality Minded are also significantly more likely than any other mindset (47 percent) to say that they believe participating in the census could benefit them personally.

Although the Confidentiality Minded mindset are concerned about what will happen with their answers to the 2020 Census, this mindset group has the largest percentage of people who trust all levels of government. Seventy-six percent of the Confidentiality Minded trust the federal government "just about always" or "most of the time." Eighty-two percent trust their state government and 84 percent trust their local government.

What are their potential motivators for participation?

The largest share of the Confidentiality Minded (31 percent) said the most important reason, to them personally, to fill out the census form was that "It helps determine funding for public services in my community like schools and fire departments." Nearly all respondents in the Confidentiality Minded mindset value hospitals and healthcare (98 percent), fire departments (97 percent), police departments (96 percent), and roads and highways (96 percent). Figure 20 reports the percentage of Confidentiality Minded who said each motivator included in the 2020 CBAMS Survey was "very important" or "extremely important" to them.

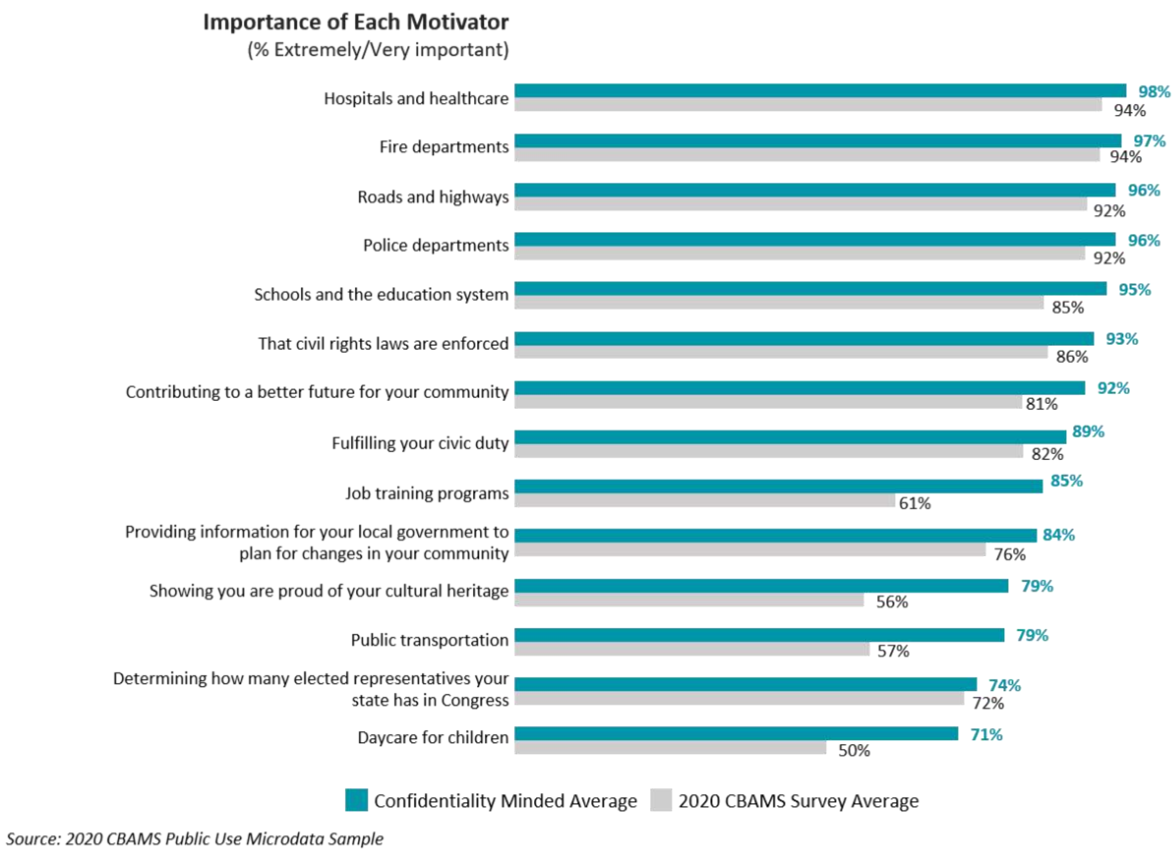


Figure 20. Confidentiality Minded Motivators

5.2.2.4 Head Nodders

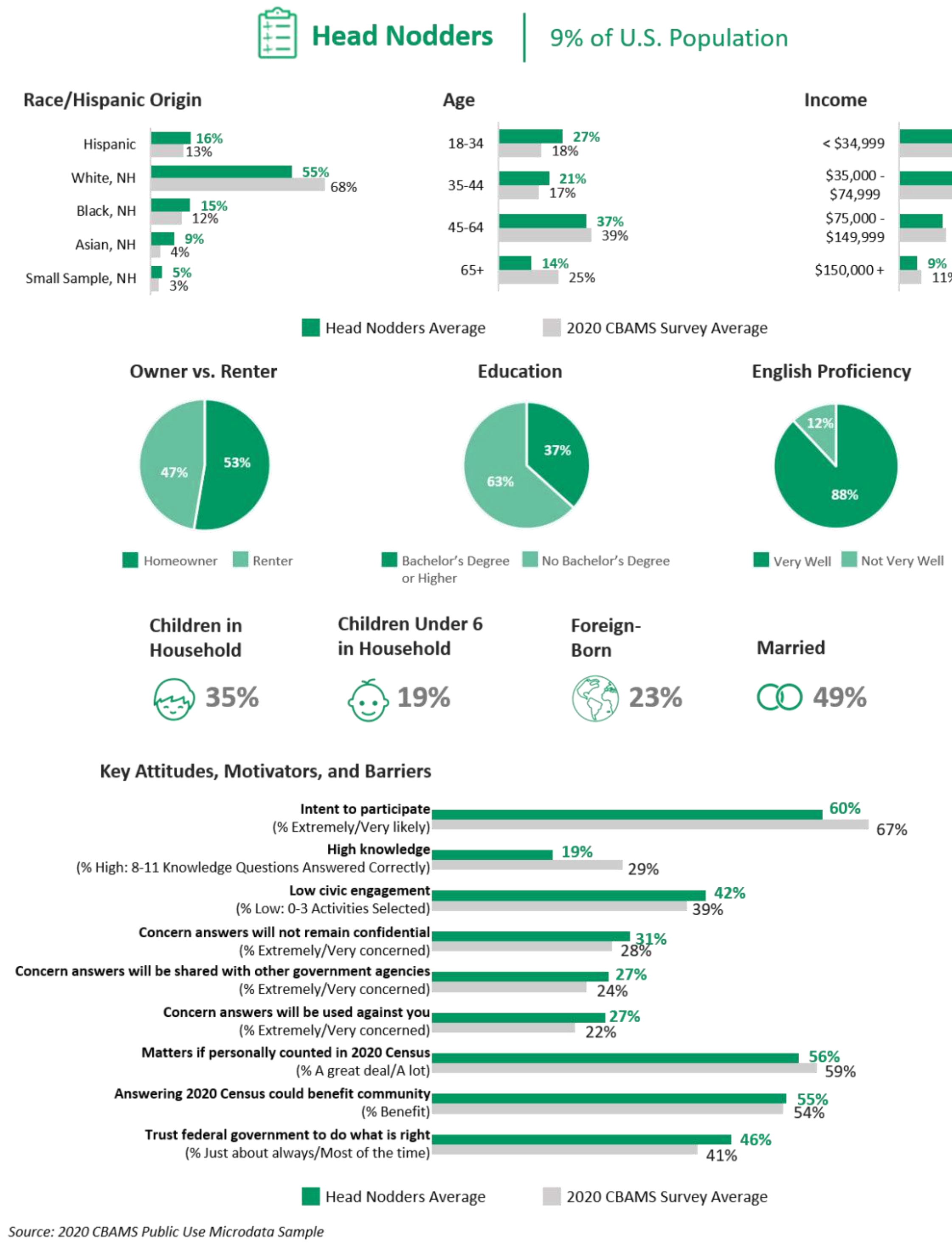


Figure 21. Head Nodders Key Characteristics

Who are they?

Nine percent of the U.S. population has a Head Nodders mindset, as reported in Figure 21. The Head Nodders mindset group is fairly diverse, with 16 percent of respondents being of Hispanic origin, 15 percent non-Hispanic Black or African American, 9 percent non-Hispanic Asian, and 5 percent from non-Hispanic small-sample race. Only 55 percent of the mindset is non-Hispanic White. Twenty-three percent of Head Nodders are foreign-born, and 12 percent speak English at a level below “very well.”

As Figure 21 also shows, 27 percent of Head Nodders are 18 to 34, making this the mindset with the largest proportion of young people. However, as with all other mindset groups except the Disconnected Doubters, the largest share of respondents are people 45 to 64 (37 percent). Those who are 35 to 44 compose 21 percent of the mindset followed by people 65 or older at 14 percent. An above-average percentage (79 percent) of Head Nodders use the internet frequently.

Members of the Head Nodders group tend to have fairly low incomes, with 35 percent earning $34,999 or less annually and 34 percent between $35,000 and $74,999. Twenty-two percent of those in this mindset group earn between $75,000 and $149,999, and 9 percent earn $150,000 or more annually. There are slightly more homeowners (53 percent) than renters (47 percent) in this group. Forty-nine percent are married, and 35 percent live in households with children age 17 or younger.

Do they intend to respond, and how do they think about the census?

Figure 21 shows that 60 percent, a slightly below-average percentage, of those with the Head Nodders mindset plan to respond to the 2020 Census. The hallmark of this mindset is its propensity to answer knowledge questions in the affirmative, regardless of whether the statement is true. Fifty-nine percent of Head Nodders are categorized as medium knowledge. However, an inspection of answers to the questions reveals that this is, in part, because respondents of the Head Nodders mindset tended to respond to knowledge questions with “yes” more often than any other mindset. This pattern of response resulted in Head Nodders getting the same knowledge questions right and wrong, depending on whether or not “yes” was the correct response.

The largest share of respondents in the Head Nodders mindset are classified as having low levels of civic engagement (42 percent). Just over three-fourths (76 percent) have voted in an election, and just over half have volunteered at any organization (55 percent) or signed a petition (51 percent).

What are their potential barriers to participation?

Members of the Head Nodders mindset group are moderately concerned about confidentiality and repercussions from their participation in the census. Figure 21 shows that 31 percent of people in the Head Nodders mindset are “extremely concerned” or “very concerned” that the Census Bureau will not keep their answers confidential, and 27 percent are concerned that the Bureau will share their answers with other government agencies. The same percentage, 27 percent, fear their responses to the 2020 Census could be used against them.

A below-average percentage (56 percent) of the Head Nodders mindset feel it matters “a great deal” or “a lot” if they are counted in the 2020 Census. A similar amount, 55 percent, believe that participating in the census could benefit their community, while 42 percent believe participation could benefit them personally.

Roughly half, and slightly more than the average percentage, of respondents in the Head Nodders mindset are trusting of the government. Forty-six percent trust the federal government to do what is right “just about always” or “most of the time,” 50 percent trust their state government, and 58 percent trust their local government.

What are their potential motivators for participation?

As with most other mindsets, respondents in the Head Nodders mindset group were most likely to say the most important reason to participate in the 2020 Census is that it “helps determine funding for public services in my community.” A high percentage of this group responded favorably to a number of potential motivators. For example, over 90 percent of respondents in the Head Nodders group said that each of the following motivators was “very important” or “extremely important”: hospitals and healthcare (97 percent), fire departments (96 percent), roads and highways (92 percent), police departments (92 percent), schools and the education system (91 percent), and that civil rights laws are enforced (90 percent). The percentage of Head Nodders who said each motivator is “very important” or “extremely important” is reported in Figure 22.

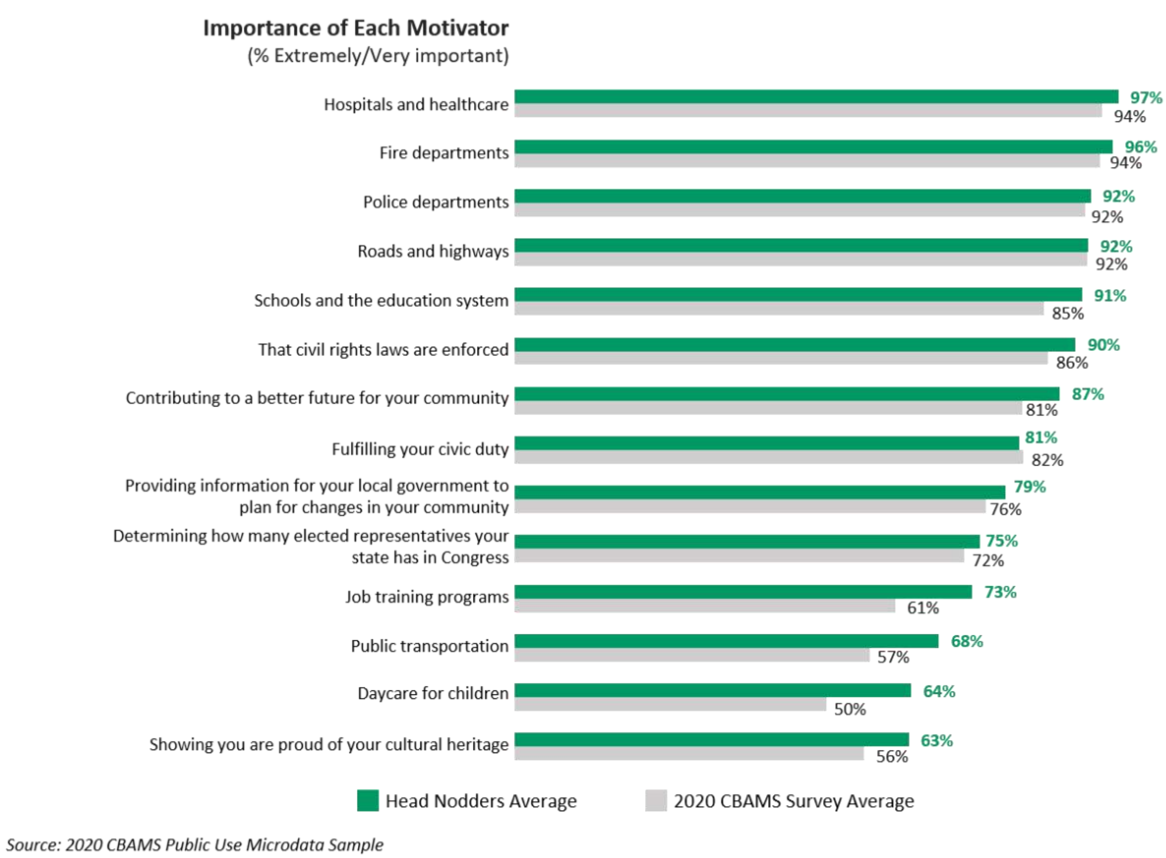


Figure 22. Head Nodders Motivators

5.2.2.5 Wary Skeptics

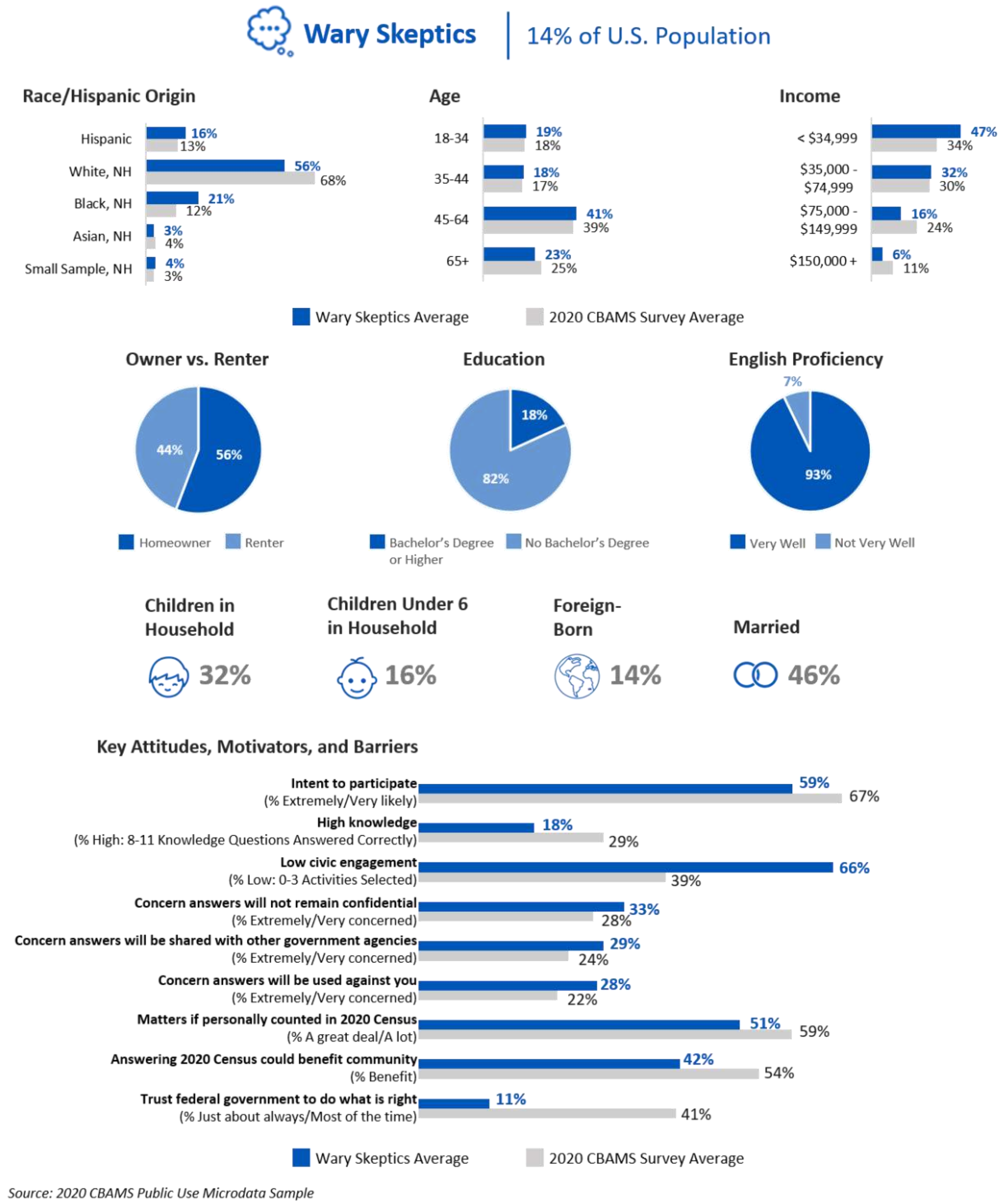


Figure 23. Wary Skeptics Key Characteristics

Who are they?

As Figure 23 shows, 14 percent of the population is in the Wary Skeptics mindset group. This mindset group has an above-average level of diversity in race and Hispanic-origin groups and has a composition of these groups similar to the Head Nodders mindset, as reported in Figure 23 and Figure 21. Fifty-six percent of Wary Skeptics are non-Hispanic White, 21 percent are non-Hispanic Black or African American, 16 percent are of Hispanic origin, 4 percent are non-Hispanic small-sample race, and 3 percent are non-Hispanic Asian.

Wary Skeptics skew slightly older, with 64 percent of respondents 45 or older. Eighteen percent are 35 to 44, and 19 percent are 18 to 34.

Nearly half (47 percent) earn $34,999 or less annually, and 32 percent earn between $35,000 and $74,000. More than half (56 percent) of Wary Skeptics own their home, and 44 percent are renters.

Less than half (46 percent) of Wary Skeptics are married, and 32 percent live in households with children 17 or younger. Wary Skeptics have a below-average frequency of internet use, with 62 percent using the internet “almost constantly” or “several times a day.” A small percentage (14 percent) are foreign-born, and only 7 percent speak English at a level below “very well.”

Do they intend to respond, and how do they think about the census?

Figure 23 shows that a below-average percentage (59 percent) of Wary Skeptics say they are “very likely” or “extremely likely” to respond to the 2020 Census. They have low levels of knowledge about the census, with almost half of Wary Skeptics categorized as “low knowledge,” meaning they answered a maximum of four knowledge questions correctly. Only 36 percent of Wary Skeptics know the census is used to decide how much money communities will receive from the government compared with the average 45 percent across all mindset groups.

The majority of those in the Wary Skeptics mindset also have low levels of civic engagement (66 percent). A mere 5 percent are considered to have high levels of civic engagement. Only 67 percent of the Wary Skeptics group report voting in an election, compared with the average of 82 percent across all mindsets.

What are their potential barriers to participation?

Figure 23 shows that a slightly above-average percentage of the Wary Skeptics mindset group is concerned that participating in the census could yield negative consequences. Thirty-three percent are “very concerned” or “extremely concerned” that the Census Bureau will not keep their answers to the 2020 Census confidential, 29 percent are concerned that the Census Bureau will share their answers with other government agencies, and 28 percent fear their answers could be used against them.

Wary Skeptics, along with Disconnected Doubters, discussed in the next section, have the lowest share of respondents (51 percent) who believe that it matters “a great deal” or “a lot” if they are personally counted in the 2020 Census. This perception that participation matters very little is consistent with the low percentages of Wary Skeptics who think it could benefit their community (42 percent) or them personally (30 percent) if they participate.

Wary Skeptics also have significantly lower levels of trust in the government than all other mindsets. Only 11 percent trust the federal government to do what is right “just about always” or “most of the time.” Ten percent trust their state government, and 14 percent trust their local government “just about always” or “most of the time.”

What are their potential motivators for participation?

Members of the Wary Skeptics group are just as likely as other mindsets (32 percent) to identify “It helps determine funding for public services in my community like schools and fire departments” as the most important reason to participate in the census. Another 22 percent cited “civic duty” as the most important reason to participate in the census.

However, lower percentages of Wary Skeptics rate each potential motivator as “very important” or “extremely important.” For instance, Figure 24 shows that support for not one single motivator exceeds 86 percent of this mindset group. Similar to other mindsets, the motivators that Wary Skeptics do tend to cite are roads and highways (86 percent), hospitals and healthcare (85 percent), and civic duty (85 percent) as their most valued motivators.

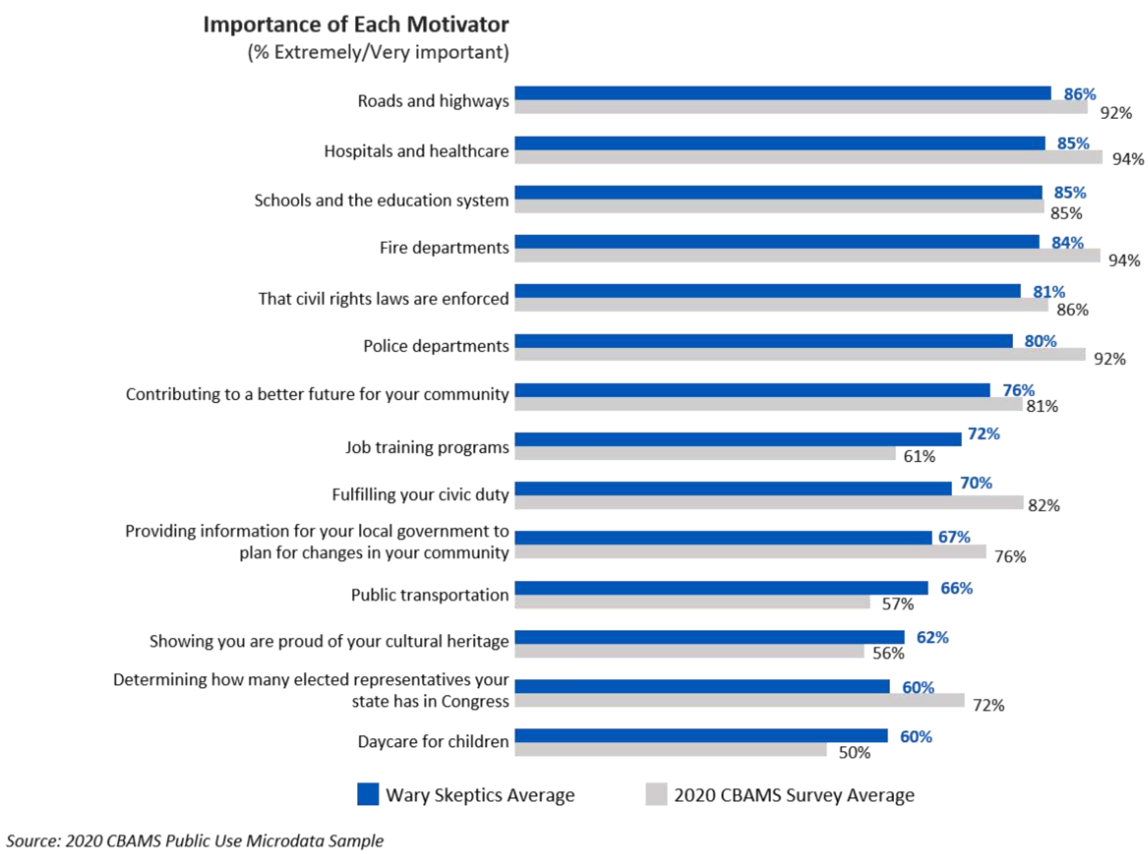


Figure 24. Wary Skeptics Motivators

5.2.2.6 Disconnected Doubters

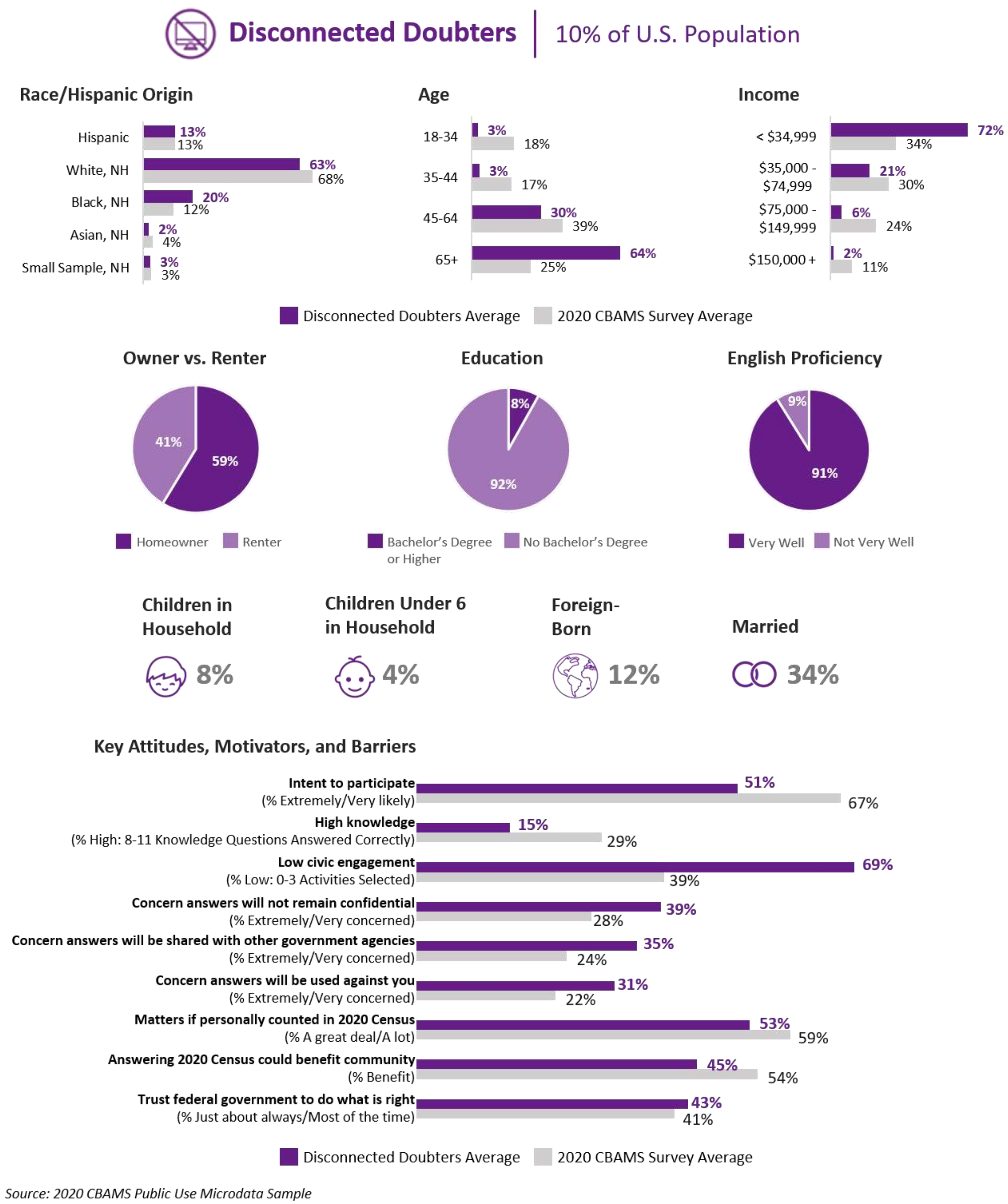


Figure 25. Disconnected Doubters Key Characteristics

Who are they?

Ten percent of the population is part of the Disconnected Doubters mindset group (see Figure 25), which has a slightly above-average level of diversity. Altogether, 63 percent of Disconnected Doubters are non-Hispanic White, 20 percent are non-Hispanic Black or African American, 13 percent are Hispanic, 3 percent are non-Hispanic small-sample race, and 2 percent are non-Hispanic Asian.

A majority of Disconnected Doubters are 65 or older (64 percent) and another 30 percent are between 45 and 64. A mere 6 percent of Disconnected Doubters are under 35. Very few use the internet with any real frequency, with only 6 percent using the internet “almost constantly” or “several times a day.”

As Figure 25 shows, many members of the Disconnected Doubters mindset earn below $35,000. Nearly two-thirds (72 percent) of this mindset group earn $34,999 or less annually, and 21 percent earn between $35,000 and $74,999. Fifty-nine percent own their homes, while 41 percent are renters. Very few—only 8 percent—of Disconnected Doubters live in households with children, and only 34 percent are married. Twelve percent are foreign-born, and 9 percent speak English at a level below “very well.”

Do they intend to respond, and how do they think about the census?

Figure 25 shows that the Disconnected Doubters mindset group has the lowest percentage of people who say they intend to respond to the census (51 percent). Over half the respondents in this mindset are classified as having low knowledge of the census (54 percent). Only 56 percent of Disconnected Doubters know the census is used to determine changes that have taken place in the size, location, and characteristics of people in the United States. For comparison, in the aggregate, 80 percent of survey respondents know the census is used this way.

The majority of those in the Disconnected Doubters mindset also have low levels of civic engagement (69 percent). Although 74 percent of Disconnected Doubters have voted in an election, very few members of this group have engaged in other political activities. For example, only roughly one-quarter have signed a petition (24 percent) or volunteered at an organization (25 percent).

What are their potential barriers to participation?

Those of the Disconnected Doubters mindset have above-average levels of concern about the confidentiality and use of their responses to the census. Figure 25 shows that 39 percent are concerned the Census Bureau will not keep their answers to the 2020 Census confidential, 35 percent are concerned the Census Bureau will share their answers with other government agencies, and 31 percent are concerned their answers to the census will be used against them.

Similarly, only 53 percent believe it matters if they personally are counted in the 2020 Census. Very few perceive community (45 percent) or personal (34 percent) benefits from participating in the census. And 3 percent—the highest of any mindset group—believe participating in the census could harm them personally.

Although Disconnected Doubters tend to exhibit a number of attitudes and behaviors inconsistent with participation in the census, they have levels of trust in the federal government (43 percent) close to the survey average of 41 percent.

What are their potential motivators for participation?

When asked to identify the single most important reason to participate in the census, 30 percent of Disconnected Doubters selected "It helps determine funding for public services in my community like schools and fire departments." Another 25 percent said that "civic duty" was the most important reason they personally should fill out the census form.

Figure 26 shows that members of the Disconnected Doubters group value some public services at a fairly high rate. Roughly 90 percent said that fire departments (92 percent), hospitals and healthcare (91 percent), or police departments (90 percent) were "very important" or "extremely important" to them.

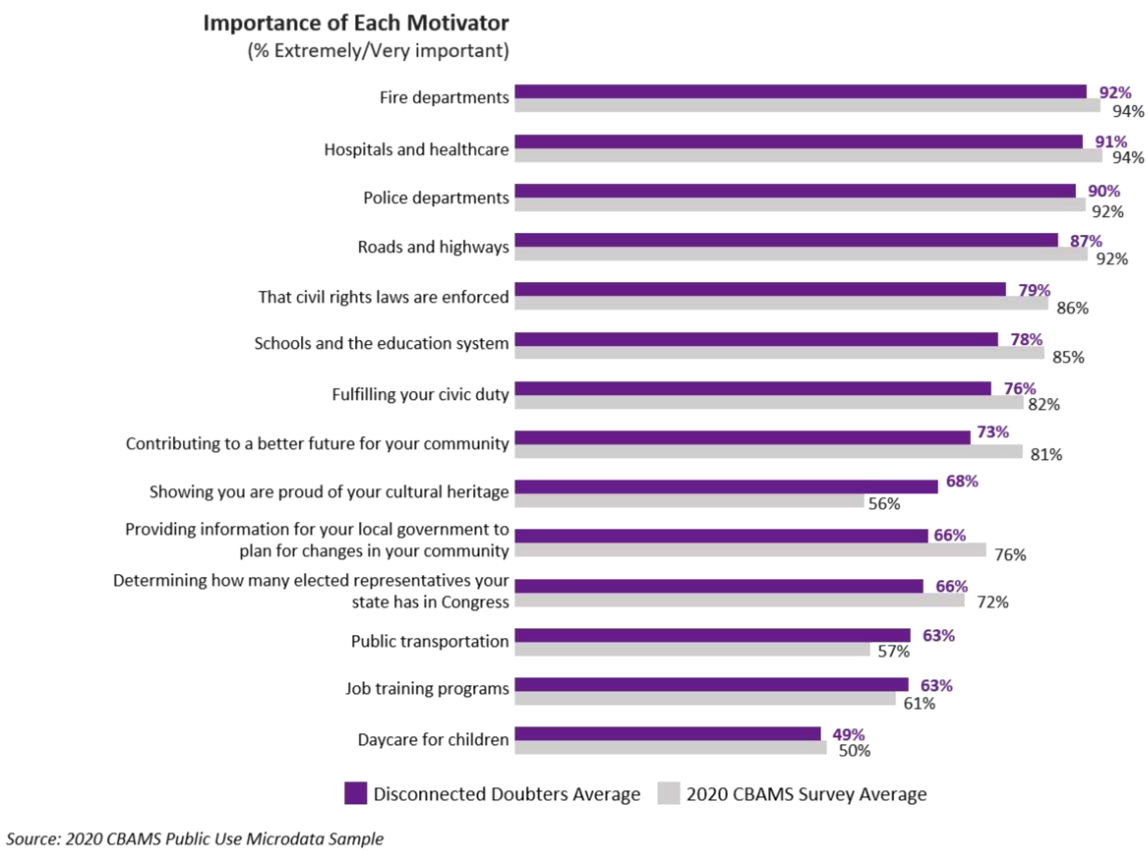


Figure 26. Disconnected Doubters Motivators

6. Tract Segmentation

6.1 Methodology

6.1.1 Data Sources

The CRAT team used multiple data sources to create a set of candidate tract segmentations from which the final segmentation solution was selected (see Table 8). Three of these data sources—the 2020 Predicted Self-Response Score (2020-PSRS), the 2020 Internet Proportion of Self-Response (2020-IPSR), and the mindset composition of segments—were produced through research described in previous sections.

Data on the media consumption patterns of people living in census tracts came from MRI's *Survey of the American Consumer® Doublebase 2018* data set, which provides information on media usage by segment for all candidate segmentations. All other data sources came from Census Bureau data collections. The variables derived from these data sources include the following: tract-level demographic variables from the 2016 PDB that were used to create the 2020-PSRS; a measure of tract-level urban density published by the Census Bureau's Geography Division; two tract-level internet access variables calculated from ACS responses from 2013 through 2016; and a measure of the timing of self-response to the 2010 Census (i.e., the number of days it took each tract to reach a 50 percent participation rate) calculated using publicly available participation rate data.

The CRAT team evaluated potential segments using mindset estimates and media usage data at the segment level rather than using these data sources to develop potential segmentations. All other data sources served as inputs for creating potential tract segmentations.

Table 8. Tract-Level Segmentation Data Source Usage

Segmentation Creation and Evaluation	Segmentation Evaluation Only
2020 Predicted Self-Response Score (PSRS)	Media usage data from MRI's Survey of the American Consumer®, Doublebase 2018
2020 Internet Proportion of Self-Response (IPSR)	Segment-level mindset estimates
Day of the 2010 Census enumeration that a tract reached 50% participation	
Variables used to create the 2020-PSRS and 2020-IPSR except the Esri Tapestry Segmentation (see *Table 9* and *Table 10* in Appendix B)	

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/,
2016 Census Planning Database. https://www.census.gov/research/data/planning_database/2016/,
Survey of the American Consumer (Doublebase 2018). https://mri.gfk.com/solutions/the-survey-of-the-american-consumerr/the-survey-of-the-american-consumerr/,
2020 CBAMS Public Use Microdata Sample,
2010 Census Take 10 Data.

6.1.2 Data Preparation

Tract-level segmentation requires complete data (i.e., no missing values). Therefore, the CRAT team used a county median imputation approach to replace missing values caused by the absence of estimators on the PDB or the absence of publicly available self-response rates to the 2010 Census. This imputation method replaced missing values with the median value of the other tracts belonging to the same county. The CRAT team used a factor analysis with varimax rotations to reduce the number of potential inputs to tract segmentation into a smaller number of components that had strong associations with individual input variables. The factor scores for each tract on these seven varimax-rotated factors provided the inputs into the k-means and Ward's method algorithms.[26] Technical Appendix B.5 describes additional details on the tract segmentation variables and their loadings.

6.1.3 Creating Candidate Tract Segmentations

To generate and select the final tract segmentation solution, the CRAT team used a process similar to what was used to produce a final mindset solution. Tract segments were created using k-means and Ward's method segmentation methods (described in detail in Technical Appendix B.6) to identify similar groups. However, where the mindset solutions were groupings of similar survey respondents, the tract segmentation grouped similar census tracts based on predicted self-response patterns and demographic characteristics.

6.1.4 Distributing Mindsets Across Tract-Level Segments

To create segment-level estimates of the distribution of mindsets in each segment, the CRAT team matched all CBAMS respondents to segments and then constructed segment-level estimates using appropriate demographic weights. Because each CBAMS respondent had an assigned mindset, it was possible to link mindsets to each of the potential segments.

The CRAT team weighted the portion of the CBAMS sample that fell within each potential segment to mirror the demographic distribution of that segment. ACS five-year tract-level demographic estimates aggregated using population weights served as marginal totals for each potential segment. The CRAT team then used iterative proportional fitting to assign a weight to each CBAMS respondent such that the aggregated weights within each potential segment matched the corresponding total for that potential segment. Once the weights accurately reflected the potential segment's demographic distribution, the weighted average of the mindsets of respondents living in the segment produced the segment-level estimate of the distribution of mindsets for that potential segment.

6.1.5 Understanding Media Consumption Patterns of Tract-Level Segments

Mindset and tract segmentation findings are more useful to the communications campaign when paired with media consumption data, allowing communications professionals to

[26] To determine the desirable number of factors, the CRAT team performed a PCA and examined a scree plot of eigenvalues to identify that seven factors were optimal.

determine where and using which media to communicate certain messages. For instance, if the communications campaign wanted to assuage concerns that answers to the census will not be kept private, they would identify the segment of census tracts that included the largest percentage of people with the Confidentiality Minded mindset, and, based on the media consumption habits of those segments, would broadcast messaging most likely to reach people of that mindset.

The CRAT team partnered with the media and consumer research firm MRI to pair data on media consumption patterns from MRI's *Survey of the American Consumer® Doublebase 2018* data set with the final tract segments. The CRAT team sent MRI lists detailing the tracts within each candidate segment. MRI used address information for the respondents to Survey of the American Consumer to identify which candidate segments each respondent lived in. Next, the CRAT team provided MRI with the weighting targets used for the mindsets. MRI then used these targets to produce custom weights for each candidate segment and produced estimates of media consumption weighted to match the demographic characteristics of each candidate segment. MRI provided these results to the CRAT team, which Section 6.2 discusses in detail.

6.1.6 Selecting the Final Tract Segment Solution

The CRAT team selected three candidate tract segmentations and presented them to ICC creative and strategy stakeholders, who selected a final tract segmentation. All three candidate tract segmentations provided differentiation on the 2020 Predicted Self-Response Score (2020-PSRS) across segments and reflected groupings of census tracts that were similar in distribution of segment size. The three candidate solutions presented to stakeholders were:

- Solution A, which used k-means to produce eight tract segments
- Solution B, which used k-means to produce seven tract segments
- Solution C, which used Ward's method to produce eight tract segments

Team Y&R's multicultural partners and communications experts unanimously preferred Solution A to the other solutions. There was a strong consensus that Solution A provided more differentiation in the 2020-PSRS across the segments than other solutions, produced distinct differences between its two rural segments, and reflected media consumption habits and mindsets that best aligned with segment characteristics. As with selection of the final mindset solution, stakeholders generally felt Solution A provided the optimal level of detail needed to craft a successful marketing campaign.

6.2 Results

6.2.1 Introduction to Tract Segments

After the selection of tract segmentation Solution A (as discussed in Section 6.1), the research team developed in-depth profiles of each tract segment, largely using available tract-level data from the PDB. The following four questions guided profile development:

- **What proportion of the households in the segment will respond to the 2020 Census?** The CRAT team reports the average 2020-PSRS (discussed in Section 3.2.1) for each segment.

- **Who are they?** The focus is on demographic characteristics of each tract segment including age, race and Hispanic origin, education, residence type, internet access,[27] marital status, and percentage of households with children.
- **What are their media consumption patterns?** Measures of media usage and consumer behavior from MRI's Survey of the American Consumer are used to understand the media consumption patterns of each segment (newspapers: the number read in a 28-day period, magazines: the number of issues read in a month, out-of-home: the number of miles driven in a seven-day period, radio: the number of half-hours of radio listened to in a week, television: the number of half-hours watched in a week, and internet: the number of hours used in a week).
- **What is the distribution of mindsets in tract segments?** CBAMS respondents live in different tract segments, and the distribution of their mindsets determines the mindset composition of each tract segment.

6.2.2 Geographic Distribution of Tract Segments

Figure 27 shows where each segment can be found in the United States, as well as those areas that were not assigned a segment due to being in nonmailout ACS areas. The map most clearly shows rural tracts covering large areas. The Downtown Dynamic segment is located almost entirely in densely populated urban centers, such as New York City, and, therefore, is very hard to see at a national level. Student and Military Communities includes a few large geographic tracts in Southern California that are visible at a national level but cannot be seen otherwise. Visually, the two rural segments, Country Roads and Sparse Spaces, dominate the national view of tract segments. Country Roads predominates in the eastern United States, while Sparse Spaces is primarily in the western United States.

The other segments include smaller tracts that are not as visually obvious at a national level. Main Street Middle is located in tracts that contain small towns in rural areas as well as in the exurbs of larger cities and—at a national level—can be seen in some areas around Pittsburgh. Responsive Suburbia can be seen in the Northeast corridor of the United States as well as in concentrations around large cities like Chicago, Minneapolis, San Francisco, Denver, and Atlanta. Multicultural Mosaic tracts are easiest to see in a concentration in the Central Valley of California and the southern parts of Texas and New Mexico. Rural Delta and Urban Enclaves tracts are easiest to see in tracts along the Mississippi Delta and in concentrations in the southeastern United States.

[27] The CRAT team calculated internet access measures using the tract-level estimates of internet access produced as part of the modeling efforts described in this report. Responses to the 2013-2016 ACS provided the source for these estimates, as described in Table 11. The CRAT team noise injected these estimates following the differential privacy procedures described in Section 3.1.1.4.

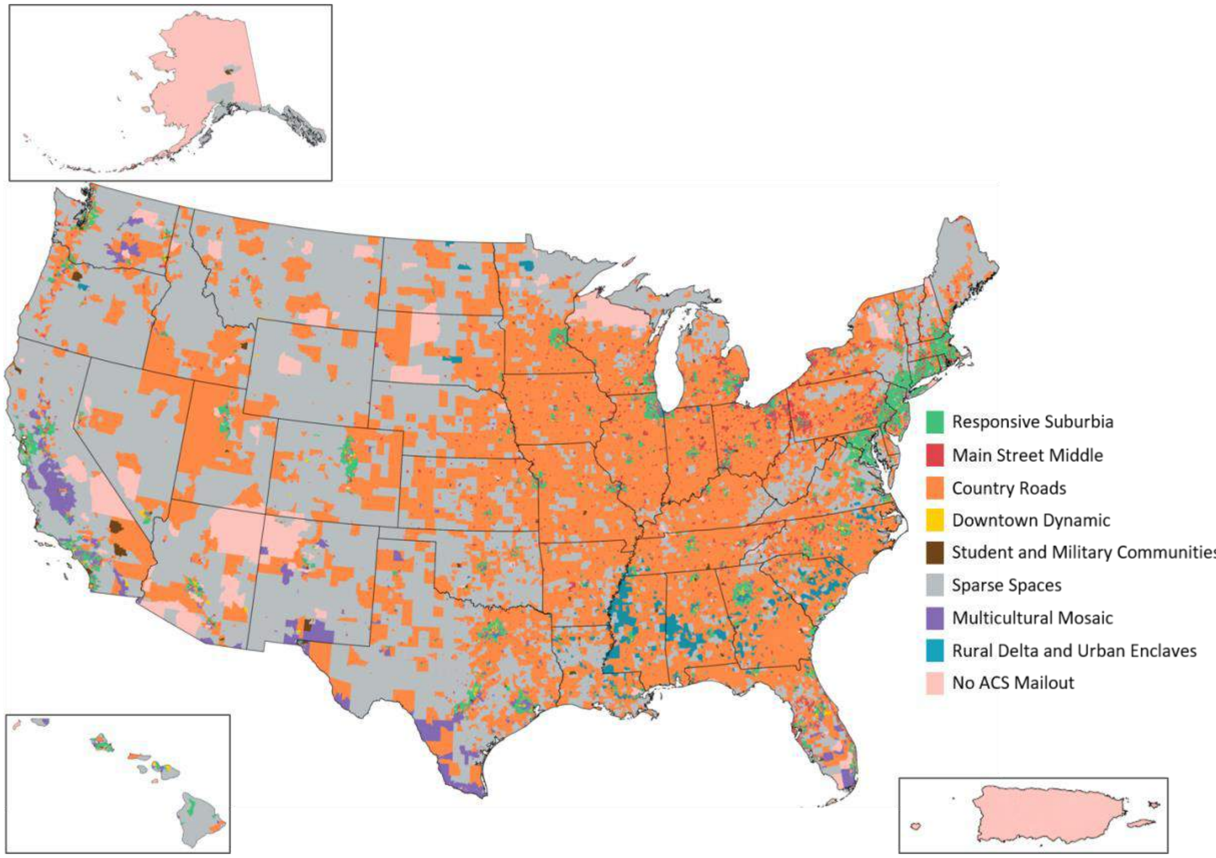


Figure 27. National View of Tract Segments

As an example of how tract segments appear on a smaller geographic scale, Washington, D.C. contains a number of tracts in each of the non-rural segments. As shown in Figure 28, tracts in the eastern half of the city are primarily in the Rural Delta and Urban Enclaves segment. While there are no Ethnic Enclave tracts in D.C. itself, there are tracts of this segment in Maryland, adjacent to Northeast D.C. Many of the universities in D.C., as well as Joint Base Anacostia-Bolling and the Pentagon in Virginia, are in tracts in the Student and Military Communities segment. Tracts on the border of Northwest D.C., as well as adjacent tracts in Virginia and Maryland, are in Responsive Suburbia. Tracts that contain the U.S. Capitol and the National Mall are in the Main Street Middle segment. The downtown parts of D.C. are included in the Downtown Dynamic segment, which covers much of Northwest D.C.

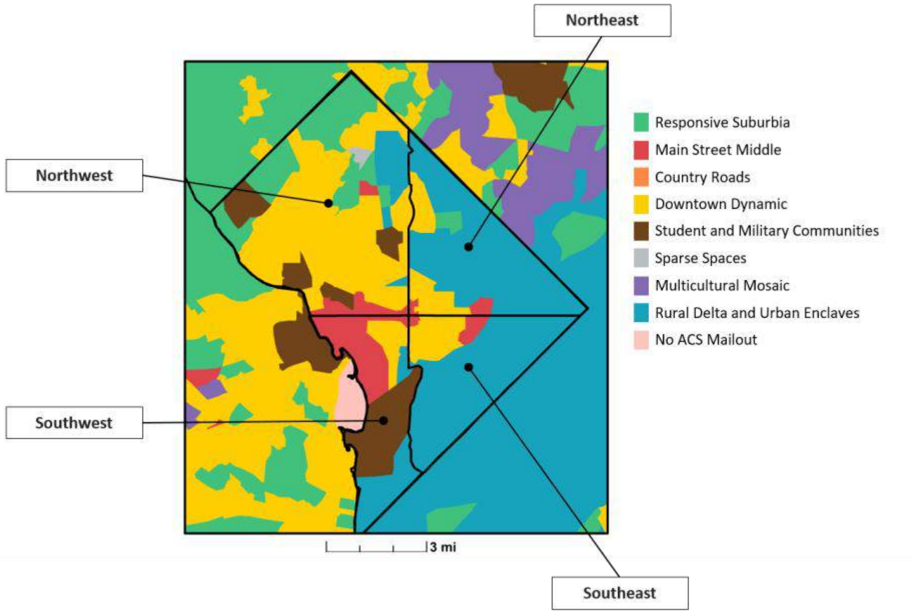


Figure 28. Washington, D.C., View of Tract Segments

6.2.3 Detailed Tract Segment Profiles

6.2.3.1 Responsive Suburbia

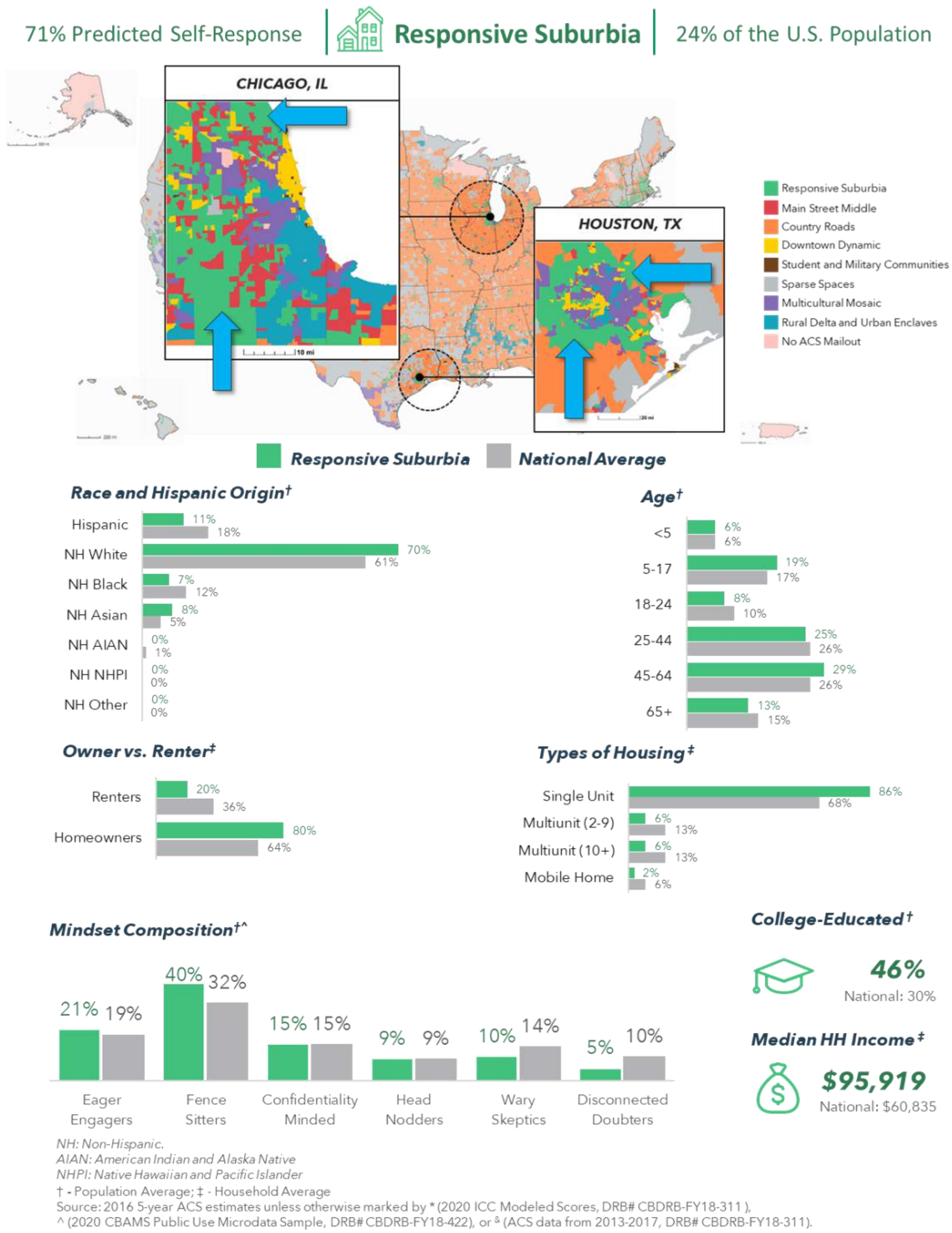


Figure 29. Responsive Suburbia

What proportion of the households in this segment will respond to the 2020 Census?

Twenty-four percent of the population lives in a Responsive Suburbia census tract (see Figure 29). Tracts classified as being in the Responsive Suburbia segment are the most likely to self-respond to the 2020 Census for both overall self-response (71 percent) and internet proportion of self-response (81 percent). This high proportion of internet self-response may be tied to the fact that 90 percent of households in the segment have access to the internet—the highest of any segment.

Who are they?

As the name suggests, Responsive Suburbia tracts are largely composed of suburban neighborhoods with single-unit homes (86 percent). Figure 29 demonstrates that the people living in this type of tract tend to have a bachelor's degree or higher (46 percent) and have a median household income of $95,919.

The Responsive Suburbia segment has the largest percentage of married couple households (63 percent) and the second-highest percentage (38 percent) of households with children 18 or younger after Multicultural Mosaic. A bit less than half of people in these tracts are 45 or older (42 percent).

Although 70 percent of people in Responsive Suburbia are non-Hispanic White, this segment is still more diverse than others (i.e., Main Street Middle, Country Roads, and Sparse Spaces) with 11 percent Hispanic, 8 percent non-Hispanic Asian, and 7 percent non-Hispanic Black or African American households.

What are their media consumption patterns?

Responsive Suburbia census tracts tend to consume media at rates fairly similar to the national population. On average, people in the Responsive Suburbia segment read newspapers and magazines, consume out-of-home media, listen to the radio, and use the internet at rates within 10 percent of the national average. However, people in the Responsive Suburbia segment view 12 percent fewer half-hours of television in one week than the national average of 64 half-hours.

What is the distribution of mindsets in tract segments?

The distribution of mindsets toward the census discussed in detail in the previous section are somewhat different in Responsive Suburbia census tracts than in the national population. Figure 29 shows that the number of Fence Sitters (40 percent) mindset members and Eager Engagers (21 percent) mindset members in the Responsive Suburbia segment is above the national average. Conversely, there are fewer Wary Skeptics (10 percent) and Disconnected Doubters (5 percent) in this segment than in the national population. This indicates that a majority of the people in Responsive Suburbia also have an above-average stated intent to respond and little concern about negative consequences from participating in the census.

6.2.3.2 Main Street Middle

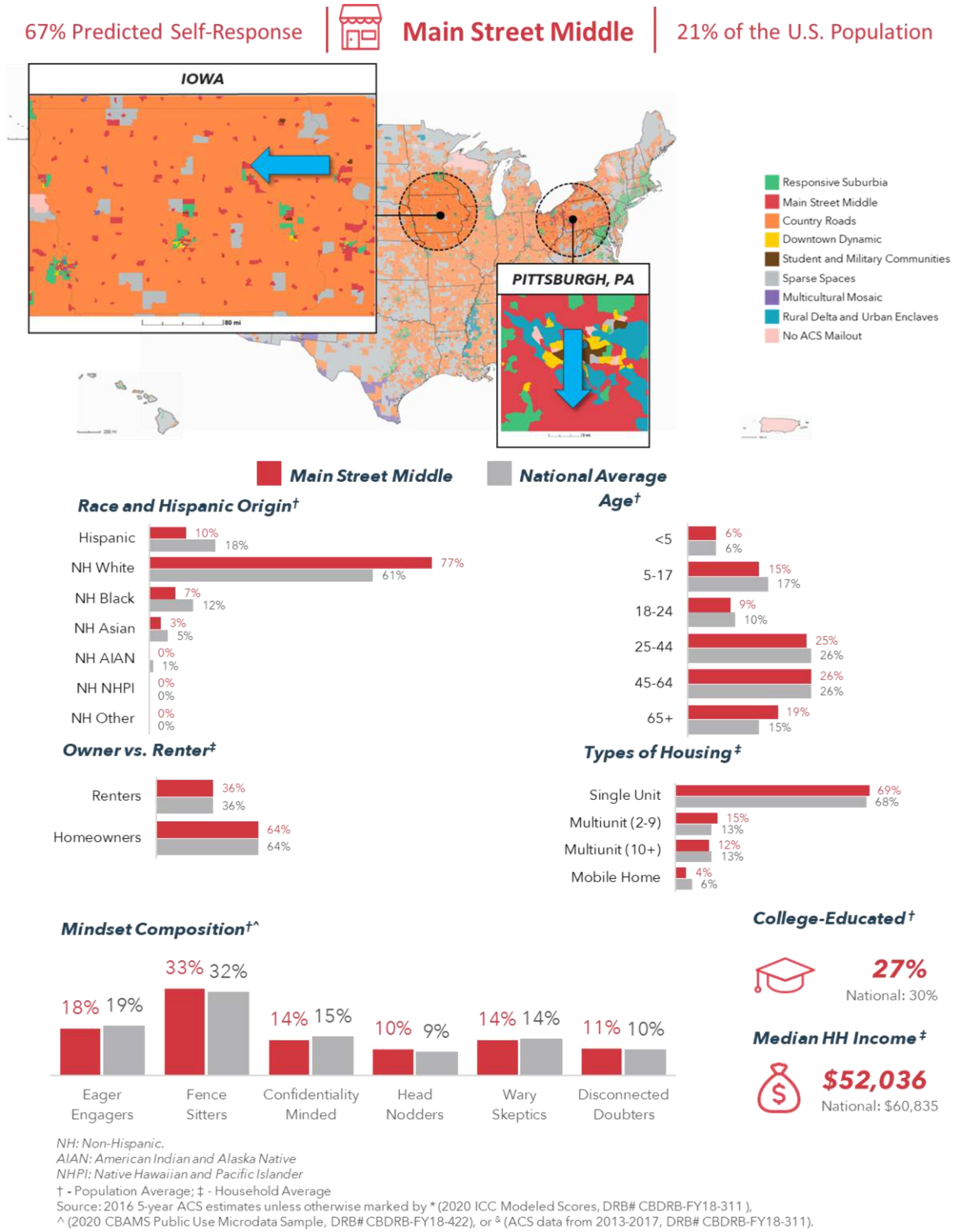


Figure 30. Main Street Middle

What proportion of the households in this segment will respond to the 2020 Census?

Tracts classified as the Main Street Middle segment are the second-most likely to self-respond to the 2020 Census (67 percent, see Figure 30). Both the predicted overall (67 percent) and internet proportion of self-response (70 percent) rates for these tracts are above average, and 78 percent of Main Street Middle households have access to the internet.

Who are they?

Census tracts classified as Main Street Middle are often made up of small towns and the less densely populated areas surrounding urban centers. Demographically, Main Street Middle tracts tend to be representative of U.S. population averages. For instance, Main Street Middle tracts have a mix of housing types: 69 percent single-unit homes, 15 percent multiunit structures ranging from two to nine units, and 12 percent multiunit structures with 10 or more units. The median household income is $52,036.

A slightly less-than-average percentage of people in Main Street Middle tracts have a bachelor's degree or higher (27 percent). Forty-four percent are married, and 28 percent have children 18 or younger living in the house.

The Main Street Middle segment has less racial and Hispanic-origin diversity than all but two other segments, Country Roads and Sparse Spaces. Figure 30 shows that 77 percent of the segment is non-Hispanic White, followed by 10 percent Hispanic, 7 percent non-Hispanic Black or African American, and 3 percent non-Hispanic Asian. This segment skews slightly older than average, with 26 percent of households 45 to 64 and 19 percent of households 65 or older.

What are their media consumption patterns?

Main Street Middle census tracts tend to consume media at rates most similar to the national population than any other segment. On average, people in the Main Street Middle segment read newspapers and magazines, consume out-of-home media, listen to the radio, view television, and use the internet at rates within 10 percent of the national average.

What is the distribution of mindsets in tract segments?

Mindsets toward the census found in the Main Street Middle segment are distributed fairly similarly to the distribution of mindsets in the national population. Figure 30 shows that there are slightly more Fence Sitters (33 percent) and Head Nodders (10 percent) and slightly fewer Eager Engagers (18 percent) and Confidentiality Minded (14 percent) in the Main Street Middle segment than in the national population. This mindset distribution is most similar to the national distribution of mindsets than any other segment. Just over one-half of the people in Main Street Middle also have an above-average stated intent to respond.

6.2.3.3 Country Roads

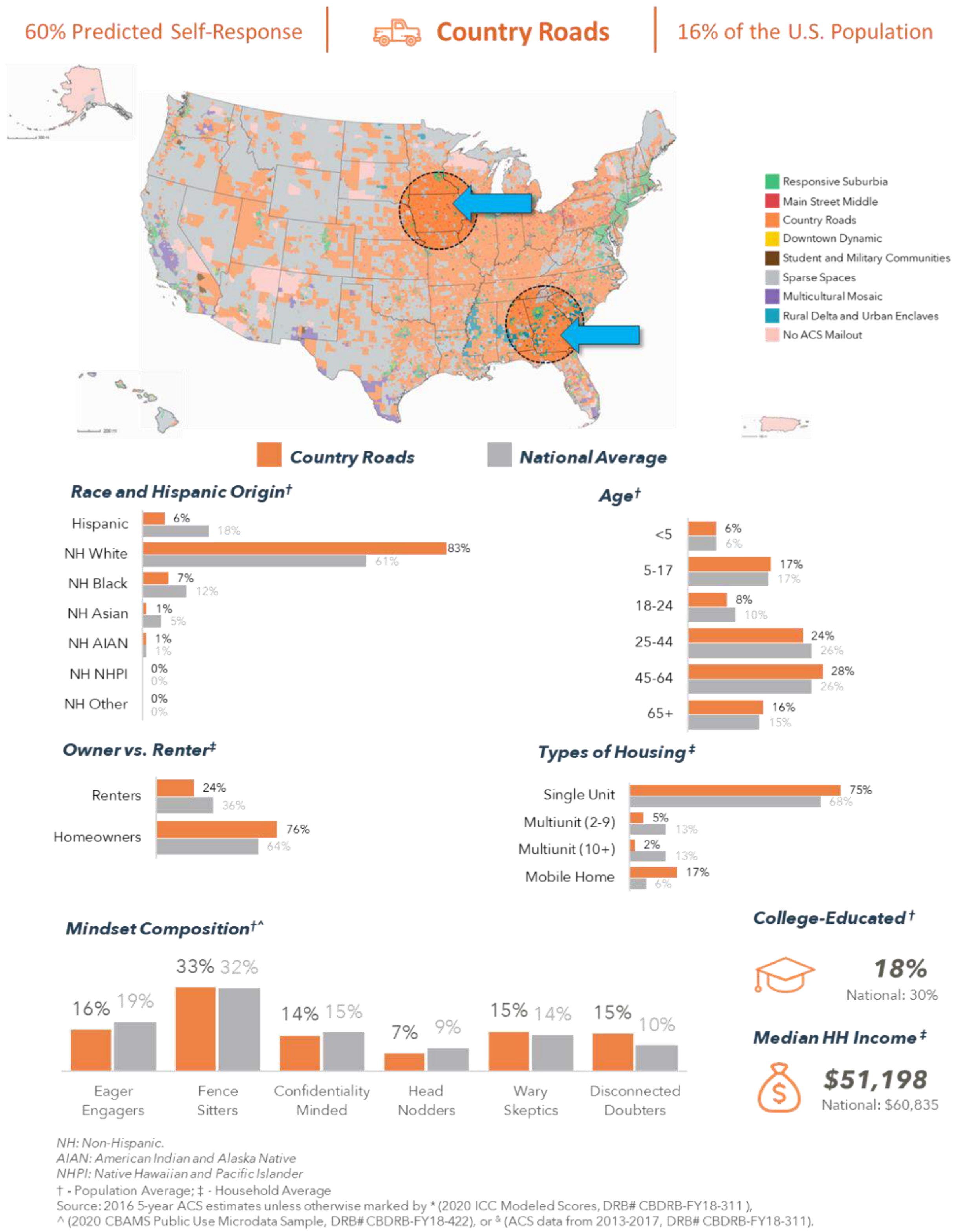


Figure 31. Country Roads

What proportion of the households in this segment will respond to the 2020 Census?

As shown in Figure 31, 16 percent of the U.S. population live in a census tract classified as Country Roads. This segment has close to an average predicted overall rate of response (60 percent), but a below-average internet proportion of self-response (57 percent). This segment also has a below-average rate of internet access (70 percent).

Who are they?

This tract segment largely encompasses people living in rural areas (69 percent). Country Roads tracts are predominantly located in the rural areas of the eastern United States and include many small towns and areas outside suburbs of major cities. Most people in Country Roads tracts own their homes (76 percent) and live in single-family homes (75 percent), but 17 percent live in mobile homes—the largest proportion of any segment. The median household income for tracts in the Country Roads segment is $51,198. Only 18 percent of people have a bachelor's degree or higher, and almost as many did not complete high school (14 percent).

Figure 31 also shows that over half of households in Country Roads tracts are married couple households (55 percent), and 32 percent of households have children 18 or younger. On average, the age distribution in these tracts is fairly similar to the distribution of age at the national level. For instance, 17 percent of people living in Country Roads tracts are 5 to 17 (the same as the national average), 28 percent are 45 to 64 (compared with 26 percent nationally), and 16 percent are 65 or older (compared with 15 nationally). Country Roads is the least diverse of all segments in its composition of race and Hispanic-origin groups, with 83 percent non-Hispanic White, 7 percent non-Hispanic Black or African American, 6 percent Hispanic, 1 percent non-Hispanic Asian, and 1 percent non-Hispanic AIAN.

What are their media consumption patterns?

Country Roads census tracts tend to consume media at rates fairly similar to the national population. On average, people in the Country Roads segment read newspapers and magazines, listen to the radio, view television, and use the internet at rates within 10 percent of the national average. However, people in the Country Roads segment consume more out-of-home media while driving or riding in a car or truck—to be exact, 17 percent more miles per week than the national average of 136 miles.

What is the distribution of mindsets in tract segments?

Like the Main Street Middle, the distribution of mindsets across the Country Roads segment is fairly similar to the distribution of mindsets in the national population. However, there are more Disconnected Doubters (15 percent) and Wary Skeptics (15 percent) present in the Country Roads segment than in the national average. There are also fewer Eager Engagers (16 percent) and Head Nodders (7 percent) than in the national population. Given the larger proportion of Disconnected Doubters and Wary Skeptics compared with the national average, this suggests that people in the Country Roads segment will be less likely to believe it matters that they are personally counted in the 2020 Census.

6.2.3.4 Downtown Dynamic

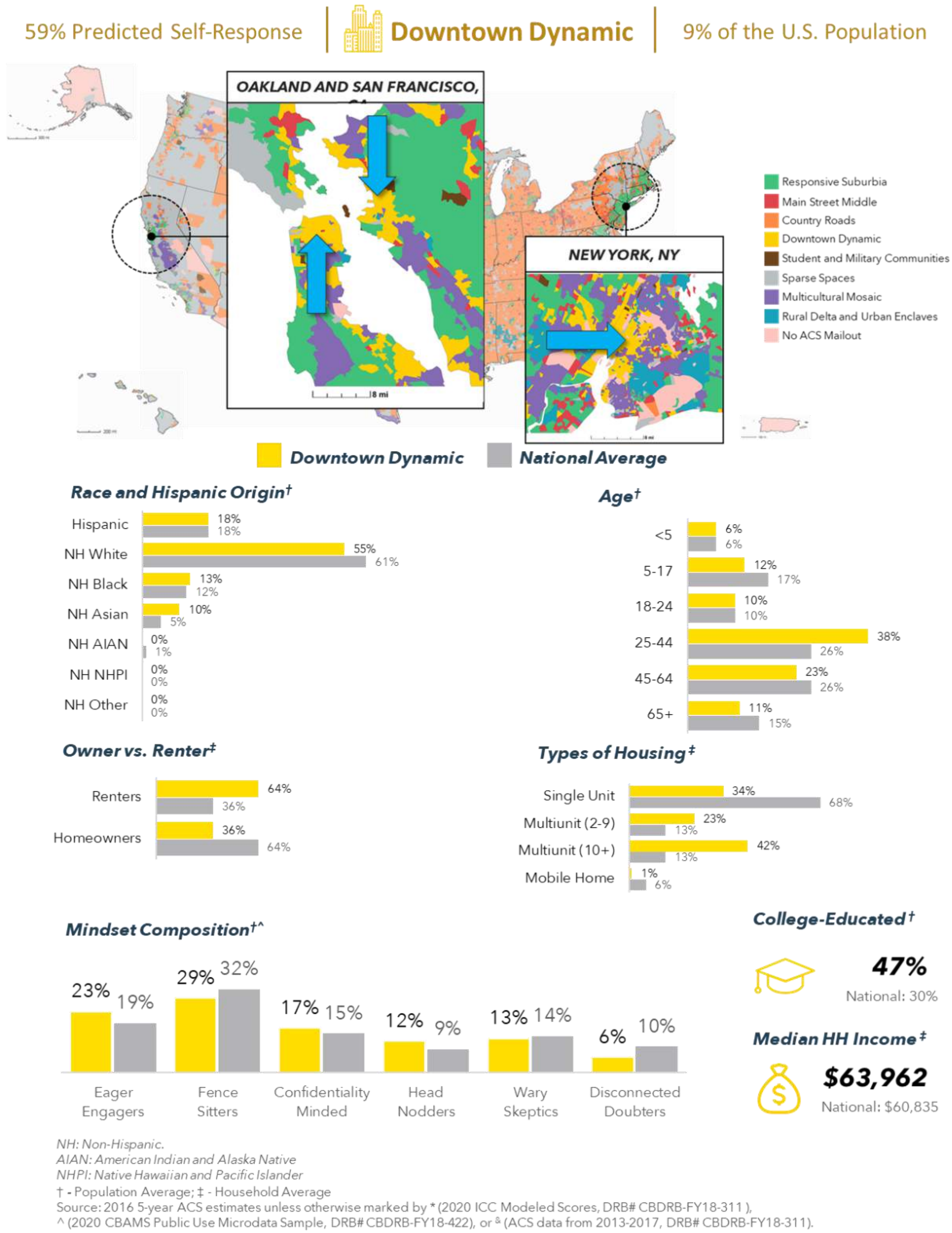


Figure 32. Downtown Dynamic

What proportion of the households in this segment will respond to the 2020 Census?

Figure 32 shows that 9 percent of the U.S. population lives in the Downtown Dynamic segment, which has 59 percent predicted overall rate of self-response and a 79 percent internet proportion of self-response. Internet access in this segment is above average, with 82 percent of households having internet access.

Who are they?

Almost all Downtown Dynamic tracts are located in densely populated metropolitan centers (98 percent), and the largest share of households are in multiunit structures made up of 10 or more units (42 percent). The median household income in this segment is just above average at $63,962, and nearly half (47 percent) of people have a bachelor's degree or higher. Only 9 percent do not have a high school diploma.

The largest share of people living in Downtown Dynamic tracts are 25 to 44 (38 percent), followed by 23 percent who are 45 to 64, and 11 percent who are 65 or older. The percentage of married couple households is below average (33 percent), and only 23 percent of households have children 18 or younger living in them, which is below the national average of 32 percent.

Figure 32 also shows that census tracts in the Downtown Dynamic segment have more racial and Hispanic-origin diversity than all but two segments, Multicultural Mosaic and Rural Delta and Urban Enclaves. Just over half of people in Downtown Dynamic tracts are non-Hispanic White (55 percent), followed by Hispanic (18 percent), non-Hispanic Black or African American (13 percent), and non-Hispanic Asian (10 percent).

What are their media consumption patterns?

Downtown Dynamic census tracts tend to consume media at rates somewhat similar to the national population. On average, people in the Downtown Dynamic segment read newspapers and magazines, listen to the radio, and view television at rates within 10 percent of the national average. However, people in the Downtown Dynamic segment consume less out-of-home media driving or riding in a car or truck (15 percent fewer miles per week than the national average of 136 miles). Additionally, people in this segment use the internet 17 percent more hours per week than the national average of 23 hours.

What is the distribution of mindsets in tract segments?

Figure 32 shows that more people in Downtown Dynamic census tracts are classified as Eager Engagers (23 percent), Head Nodders (12 percent), and Confidentiality Minded (17 percent) than the national average. Conversely, there are fewer Disconnected Doubters (6 percent), Fence Sitters (29 percent), and Wary Skeptics (13 percent). Given the larger proportion of Eager Engagers, Confidentiality Minded, and Head Nodders compared with the national average, this suggests that people in the Downtown Dynamic segment will be more likely to believe it matters that they are personally counted in the 2020 Census and that completing the 2020 Census form could personally benefit them.

6.2.3.5 Student and Military Communities

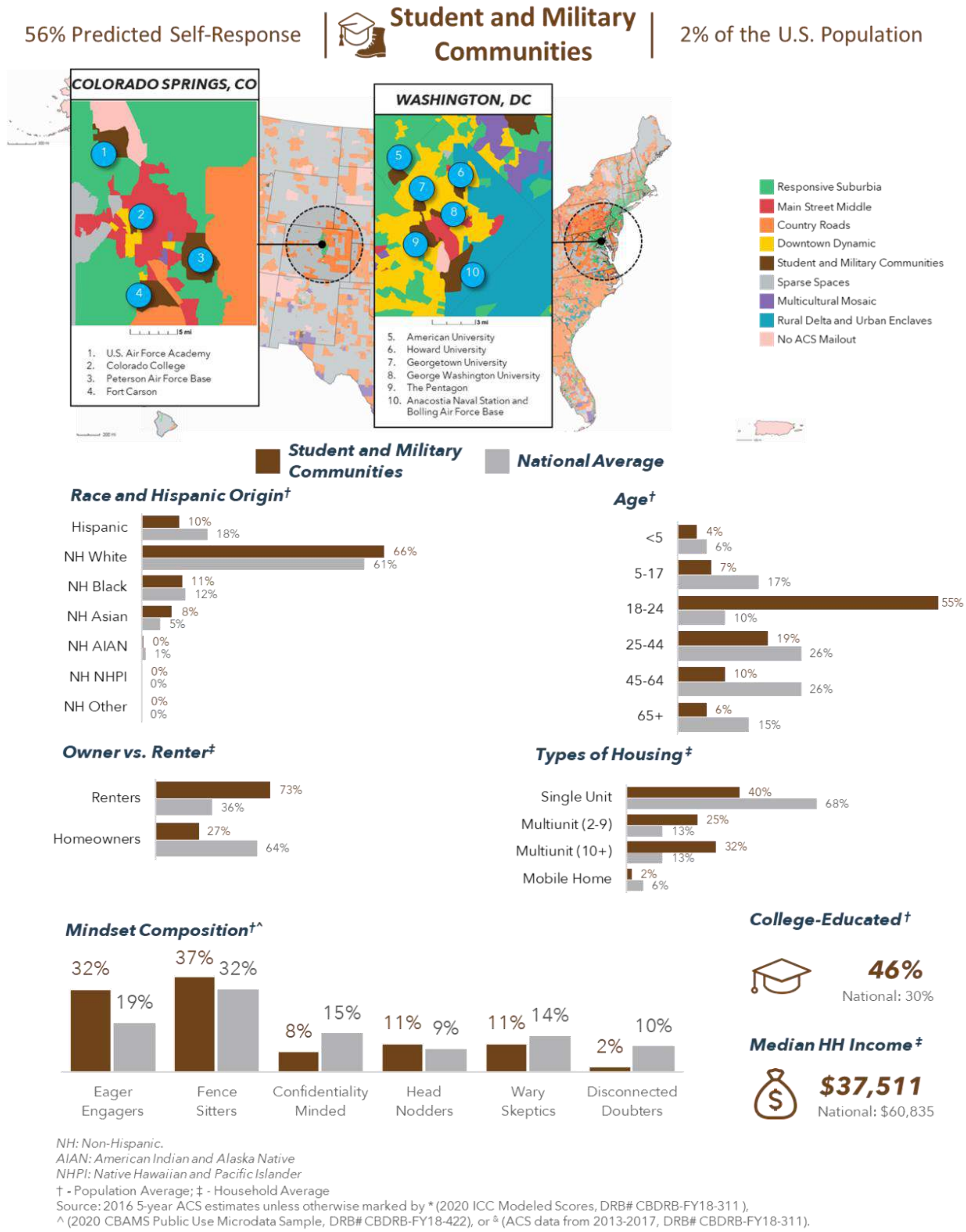


Figure 33. Student and Military Communities

What proportion of the households in this segment will respond to the 2020 Census?

Only 2 percent of the population live in a census tract classified as Student and Military Communities, as shown in Figure 33. Tracts in the Student and Military Communities segment have a below-average predicted overall rate of self-response (56 percent) but an above-average predicted proportion of internet self-response (79 percent). On average, 81 percent of the households in this segment have access to the internet.

Who are they?

As the name suggests, a large percentage of the census tracts in the Student and Military Communities segment includes noninstitutional group quarters such as dorms or barracks (32 percent). This segment has an above-average level of education, with 46 percent having earned a bachelor's degree or higher and only 8 percent that did not complete high school. The average median household income among tracts in this segment is below the national average, at $37,511.

A majority of people living in Student and Military Communities census tracts are 18 to 24 (55 percent), followed by 19 percent who are 25 to 44. Twenty-nine percent of households are occupied by married couples and only 20 percent of households include children 18 or younger—the lowest of any segment. The census tracts in the Student and Military Communities segment have slightly below-average racial and Hispanic-origin diversity, with 66 percent non-Hispanic White, 11 percent non-Hispanic Black or African American, 10 percent Hispanic, and 8 percent non-Hispanic Asian.

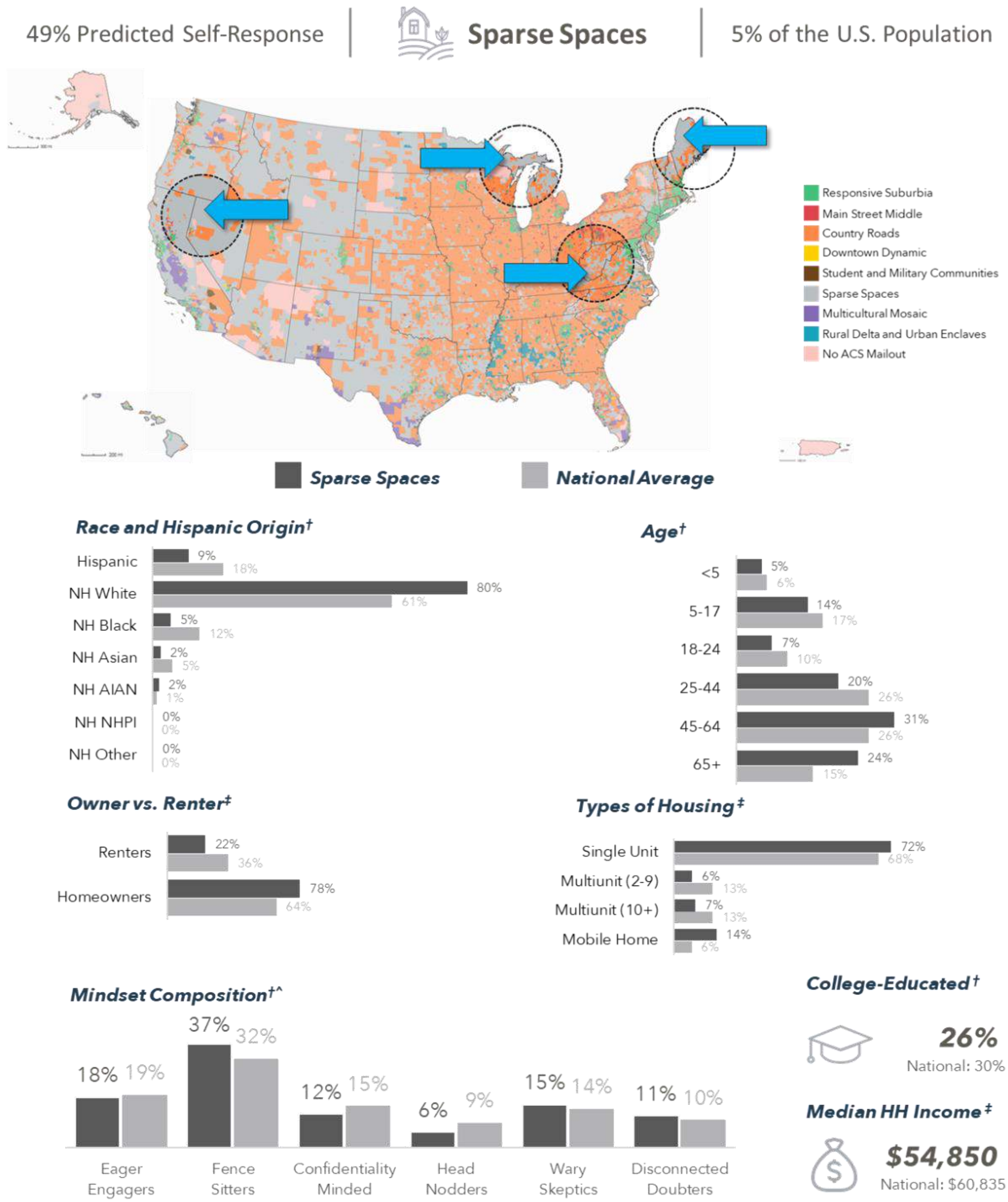
What are their media consumption patterns?

Student and Military Communities census tracts tend to consume media at lower rates than the national population for all broad categories of media, except for internet, which they consume at higher rates than average. On average, people in the Student and Military Communities segment read 16 percent fewer newspapers in a 28-day period than the national average of 15 newspapers. They also read 28 percent fewer magazine issues in one month than the national average of eight issues. In one week, this segment consumes less out-of-home media driving or riding (32 percent fewer miles), listens to 22 percent fewer half-hours of radio, and views 23 percent fewer half-hours of TV compared with the national averages of 136 miles, 35 half-hours, and 64 half-hours, respectively. On the other hand, they use the internet 31 percent more hours per week than the 23 hours per week used by the nation as a whole.

What is the distribution of mindsets in tract segments?

Figure 33 shows that there are more people of the Eager Engagers (32 percent) and Fence Sitters (37 percent) mindsets in Student and Military Communities than in the national population. Similarly, there are far fewer people who have the Disconnected Doubters (2 percent) and Confidentiality Minded (8 percent) mindsets than in the overall population. This indicates that a majority of the people in Student and Military Communities have high levels of knowledge about the census and little concern about negative consequences of participating in the census.

6.2.3.6 Sparse Spaces



NH: Non-Hispanic.
AIAN: American Indian and Alaska Native
NHPI: Native Hawaiian and Pacific Islander
† - Population Average; ‡ - Household Average
Source: 2016 5-year ACS estimates unless otherwise marked by * (2020 ICC Modeled Scores, DRB# CBDRB-FY18-311),
^ (2020 CBAMS Public Use Microdata Sample, DRB# CBDRB-FY18-422), or & (ACS data from 2013-2017, DRB# CBDRB-FY18-311).

Figure 34. Sparse Spaces

What proportion of the households in this segment will respond to the 2020 Census?

As shown in Figure 34, 5 percent of the U.S. population lives in a Sparse Spaces census tract. The predicted overall rate of self-response for tracts in this segment is below average (49 percent), and roughly 54 percent of those self-responses are predicted to be online. On average, 61 percent of households in the Sparse Spaces segment have internet access, which is below the national average.

Who are they?

Census tracts in the Sparse Spaces segment tend to be located in rural areas (68 percent), especially in the western United States, Appalachia, northern Maine, and Michigan's Upper Peninsula. Most households are occupied by their owner (78 percent) and are single-unit homes (72 percent). The average median household income in Sparse Spaces tracts is just below the national average, at $54,850. Twenty-six percent of people have a bachelor's degree or higher and 12 percent do not have a high school diploma.

Sparse Spaces census tracts are the second-least diverse, after Country Roads, with 80 percent non-Hispanic White and less than 10 percent in any other race or Hispanic-origin group.

People in this segment tend to be older than those in all other segments. For instance, 31 percent are 45 to 64, and 24 percent are 65 or older, meaning that over half of the segment population is 45 or older. Over half are also married (54 percent), but a below-average percentage live in houses with children (23 percent).

What are their media consumption patterns?

Sparse Spaces census tracts tend to consume media at rates fairly similar to the national population. On average, people in the Sparse Spaces segment read newspapers and magazines, listen to the radio, view television, and use the internet at rates within 10 percent of the national average. However, people in the Sparse Spaces segment consume more out-of-home media driving or riding in a car or truck (18 percent more miles per week than the national average of 136 miles).

What is the distribution of mindsets in tract segments?

There are more people who have the Fence Sitters (37 percent), Disconnected Doubters (11 percent), and Wary Skeptics (15 percent) mindsets in census tracts classified as Sparse Spaces than in the overall population, as shown in Figure 34. Conversely, this segment has fewer people of the Head Nodders (6 percent), Confidentiality Minded (12 percent), and Eager Engagers (18 percent) mindsets. Given the larger proportion of Fence Sitters, Wary Skeptics, and Disconnected Doubters compared with the national average, this suggests that people in the Sparse Spaces segment will be less likely to believe it matters that they are personally counted in the 2020 Census and more likely to believe that completing the census will neither benefit nor harm their community.

6.2.3.7 Multicultural Mosaic

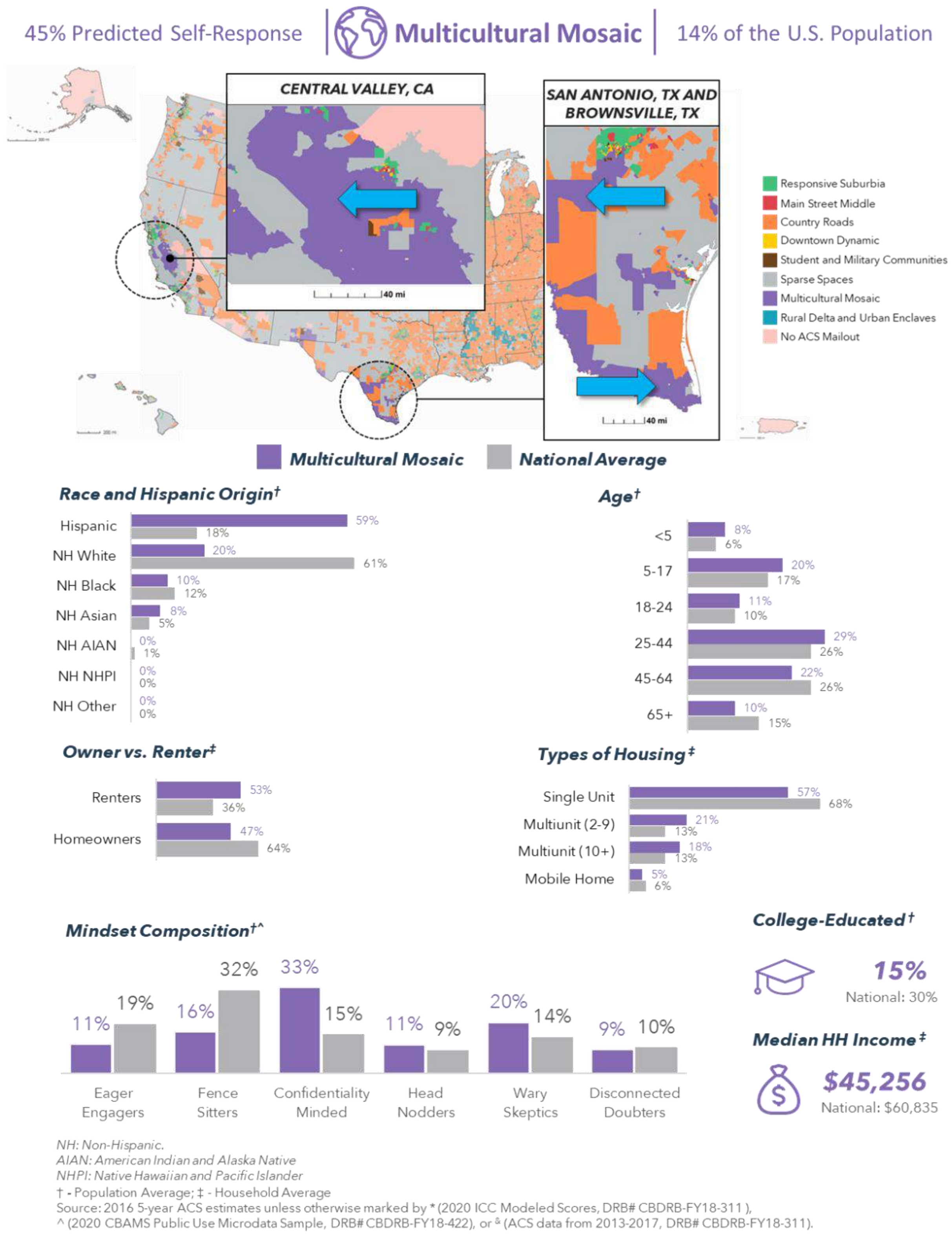


Figure 35. Multicultural Mosaic

What proportion of the households in this segment will respond to the 2020 Census?

Figure 35 reports that 14 percent of the U.S. population lives in a Multicultural Mosaic census tract. The tracts in this segment have a below-average predicted overall rate of self-response (45 percent) with a below-average expected proportion of internet self-response (56 percent). These tracts tend to be located in California's Central Valley and parts of New Mexico, Texas, and Florida, as well as concentrations in major cities.

Who are they?

As the name suggests, the Multicultural Mosaic segment is racially and ethnically diverse. A majority of people are Hispanic (59 percent), followed by 20 percent non-Hispanic White, 10 percent non-Hispanic Black or African American, and 8 percent non-Hispanic Asian. Thirty-four percent of people living in Multicultural Mosaic census tracts were born outside of the United States, well above the national average of 13 percent and more than any other segment.

Multicultural Mosaic census tracts also tend to be fairly young. This segment has the largest percentage of people ages 5 to 17 of any segment (20 percent), and the single largest age group in the segment is 25 to 44 (29 percent). Given that children and people of childbearing age are both large age groups in this segment, it is not surprising that Multicultural Mosaic tracts also have the largest percentage of households with children (43 percent) among segments. Slightly less than half (45 percent) are married couple households.

The Multicultural Mosaic segment has the third-lowest median income after Rural Delta and Urban Enclaves and Student and Military Communities, at $45,256. Education rates in this segment are below average: 15 percent have a bachelor's degree or higher, and 31 percent do not have a high school diploma—the largest percentage of people without a high school diploma of any segment. A slightly below-average percentage of households are single-unit homes (57 percent), and slightly more than average are multiunit structures.

What are their media consumption patterns?

The Multicultural Mosaic segment tend to consume media at rates similar to the national population. On average, people in the Multicultural Mosaic segment read newspapers and magazines, listen to the radio, view television, and use the internet at rates within 10 percent of the national average. However, this segment consumes less out-of-home media by driving or riding in a car or truck (13 percent fewer miles per week than the national average of 136 miles).

What is the distribution of mindsets in tract segments?

The distribution of mindsets in Multicultural Mosaic census tracts is different from the distribution in the population. There are far fewer Fence Sitters (16 percent) and Eager Engagers (11 percent) in the Multicultural Mosaic segment and slightly fewer Disconnected Doubters (9 percent) than in the national population. A large percentage (33 percent) of people in Multicultural Mosaic tracts demonstrate a Confidentiality Minded mindset, meaning they are concerned about how their answers to the 2020 Census will be used. There are also more people of the Wary Skeptics (20 percent) and Head Nodders (11 percent) mindsets in these tracts than in the national population.

6.2.3.8 Rural Delta and Urban Enclaves

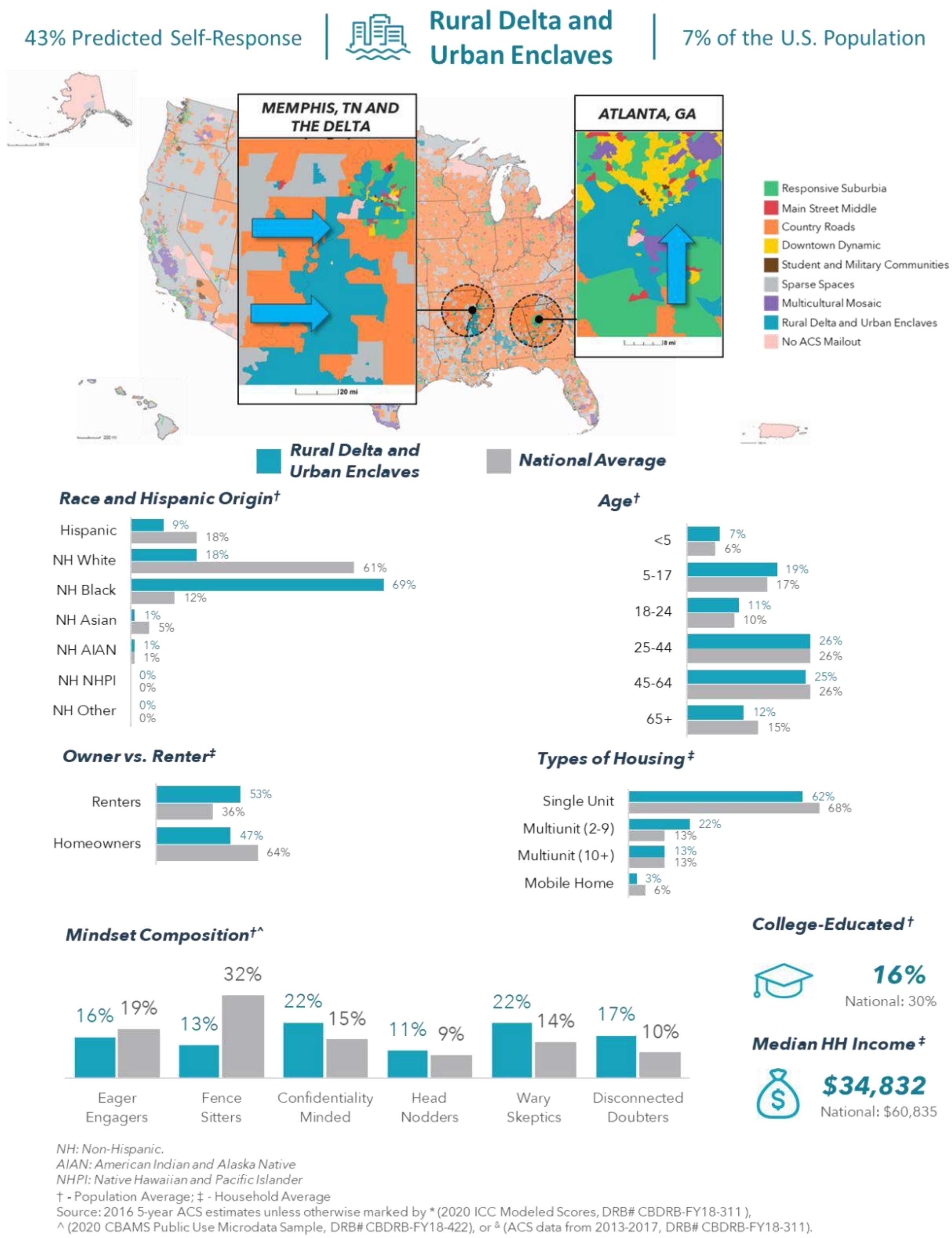


Figure 36. Rural Delta and Urban Enclaves

What proportion of the households in this segment will respond to the 2020 Census?

Tracts classified as Rural Delta and Urban Enclaves have the lowest predicted likelihood of self-responding to the census (43 percent) and contain 7 percent of the U.S. population. These tracts also have the lowest predicted proportion of internet self-response at 35 percent. Only 60 percent of tracts in this segment have internet access—the lowest of any segment.

Who are they?

Tracts in this segment can be found in rural parts of the southeastern United States, including Mississippi, Alabama, Georgia, and South Carolina, as well as concentrations in major cities. A majority of people in Rural Delta and Urban Enclaves census tracts are non-Hispanic Black or African American (69 percent). Non-Hispanic White make up 18 percent of the segment, followed by Hispanic (9 percent), non-Hispanic Asian (1 percent), and non-Hispanic AIAN (1 percent). A below-average proportion of households are in single-unit homes (62 percent), and slightly more than average are in multiunit structures with two to nine units (22 percent).

The distribution of age in this segment mirrors the distribution of age in the national population. For example, 26 percent of people in this segment are 25 to 44—the same as the national average—and 25 percent are 45 to 64, which is 1 percentage point off the national average of 26 percent. However, the percentage of married-couple households (27 percent) is far lower than the national average (48 percent). The percentage of households with children 18 or younger is 34 percent, which is just above the national average (32 percent). This segment has the lowest median household income of any segment, at $34,832. They also have below-average percentages of people with a bachelor's degree or higher (16 percent) and an above-average percentage without high school diplomas (19 percent).

What are their media consumption patterns?

Rural Delta and Urban Enclaves census tracts tend to consume media at higher rates than the national population for most broad categories of media, except for out-of-home, which they consume at lower rates than average. They do not differ more than 10 percent from the national average when it comes to the number of newspapers read. However, people in this segment read 21 percent more magazine issues in one month than the national average of eight issues. In one week, this segment listens to 14 percent more half-hours of radio, views 34 percent more half-hours of TV, and uses the internet 13 percent more hours compared with the national average of 35 half-hours, 64 half-hours, and 23 hours, respectively. On the other hand, they consume less out-of-home media by driving or riding (26 percent fewer miles per week than the 136 miles per week driven by the nation as a whole).

What is the distribution of mindsets in tract segments?

The mindsets most prevalent in this segment are the Confidentiality Minded (22 percent), Wary Skeptics (22 percent), and Disconnected Doubters (17 percent). This suggests that many people in these tracts are concerned about the confidentiality of their answers to the census and do not trust the government. There is a smaller presence of the Fence Sitters (13 percent) and Eager Engagers (16 percent) mindsets in these tracts.

7. Limitations

The major limitations of the research described in this report are as follows:

- The 17,283 respondents to the 2020 CBAMS Survey cannot provide estimates for the prevalence of mindsets in 70,000+ tracts. Therefore, the CRAT team can only link mindsets to geographic locations through estimates at the segment level. Compared with tracts, segments are larger and geographically dispersed, which results in less actionable information for the communications campaign.

- The predicted self-response scores described in this report depend on four years of ACS data from 2013 through 2016, which is less than the five years that the Census Bureau commonly uses to produce tract-level ACS aggregates. In addition, these data do not include ACS responses from 2018, the most recent year of ACS data available. Updating these results using five years of the most recent data available would improve the reliability of predictions.

- Many of the variables used to produce the predicted self-response scores described in this report derive from the 2010 Census and five-year ACS estimates from 2008-2012. By 2020, there will have been as many as 12 years between the data collection for the predictors and the behavior that they are predicting.

- U.S. Census Bureau data collections used in this report provide limited insight into the role that non-ID response and Census Questionnaire Assistance (CQA) will play during the 2020 Census. By relying primarily on the ACS to understand the 2020 Census, the CRAT team may have underestimated the amount of online non-ID response, and confusion about the internet response mode might increase the amount of response that occurs through CQA. Non-ID response might cause areas to have higher internet proportions of self-response than the 2020-IPSR predicts. Increased response through CQA could cause tracts to self-respond at rates lower than the 2020-PSRS predicts.

- This study did not account for differences in self-response caused by the differences in enumeration strategy across Type of Enumeration Areas (TEAs) that will be used during the 2020 Census. In areas that self-respond to the ACS, the data collection approach used by the Census Bureau is similar to the approach used for the decennial census in the Self-Response TEA only, so this study did not account for differences in self-response behavior that other types of enumeration might cause. Five percent of households will not be in the Self-Response TEA during the 2020 Census (Census Bureau, 2019).

- During the 2020 Census, some tracts will receive bilingual Spanish and English mail materials, which the ACS does not provide. Any differences in self-response that might be caused by the presence of these mailings in 2020 were not addressed by our modeling efforts, since our models were based on the ACS and a national-level adjustment.

8. Conclusion

This report is part of the 2020 Census Integrated Communications Campaign's (ICC) efforts to develop and execute a research-based national communications campaign to encourage participation in the 2020 Census. The 2020 Census is the first decennial census in which most people will be encouraged to respond online rather than on paper. For this reason, use of and access to the internet featured prominently in all analyses, and the CRAT team developed the 2020 Internet Proportion of Self-Response (IPSR) score to measure the predicted propensity of a census tract to self-respond to the 2020 Census via the internet.

To understand possible self-response, or unaided response, to the 2020 Census, Sections 3 and 4 of this report examined how the likelihood of self-response varies across response mode, geography, and population demographics. The CRAT team developed predictions for overall self-response and the internet proportion of self-response. It also gained a greater understanding of response and self-response timing for census tracts and demographic groups during the 2010 Census. Conclusions from these sections of the report are:

- The mode of self-response is likely to vary by contact strategy. Tracts assigned to the Internet First contact strategy will have a proportion of self-response by internet around 80 percent, while tracts assigned to the Internet Choice contact strategy will have proportions of self-response by internet around 20 percent.

- The likelihood of self-response varies across designated market areas (DMAs). The highest predicted levels of self-response to the 2020 Census in DMAs are in cities and towns in the Upper Midwest, while the lowest predicted levels of self-response are in DMAs located in the southeastern and southwestern United States. The DMAs with the most TV homes, which cover many of the country's large cities, tend to have predicted self-response rates similar to the national projected 2020 self-response rate of 60.5 percent.

- Tract-level self-response rates to the 2010 Census had an average daily interquartile range of 15 percent. The daily standard deviation of self-response at the tract-level in 2010 was never more than 1 percent away from the average of 11 percent throughout the period when participation rates were reported.

- Based on patterns observed in the timing of response to the 2010 Census, demographic groups are likely to respond at different speeds, with many of the demographic factors that affect the overall amount of self-response also influencing the speed of self-response.

Sections 5 and 6 of this report focused on understanding the distribution of views and attitudes toward the census among groupings of census tracts across the country. The CRAT team used segmentation methods to derive countrywide mindsets, or combinations of knowledge about and attitudes toward the census that are commonly shared by a group of people, and tract segments, or groups of census tracts with similar predicted self-response behavior and demographic variables associated with self-response. The CRAT team identified:

- Six key mindsets in the U.S. population that range from the highly knowledgeable and likely to participate to the disconnected and unlikely to respond. The six mindsets are:
  - Eager Engagers (19 percent of the population, 82 percent intend to respond[28])
  - Fence Sitters (32 percent of the population, 71 percent intend to respond)
  - Confidentiality Minded (15 percent of the population, 63 percent intend to respond)
  - Head Nodders (9 percent of the population, 60 percent intend to respond)
  - Wary Skeptics (14 percent of the population, 59 percent intend to respond)
  - Disconnected Doubters (10 percent of the population, 51 percent intend to respond)
- Eight tract segments in the U.S. population that range from high to low predicted self-response rate scores. The eight tract level segments are:
  - Responsive Suburbia (24 percent of the population, 71 percent predicted self-response score)
  - Main Street Middle (21 percent of the population, 67 percent predicted self-response score)
  - Country Roads (16 percent of the population, 60 percent predicted self-response score)
  - Downtown Dynamic (9 percent of the population, 59 percent predicted self-response score)
  - Student and Military Communities (2 percent of the population, 56 percent predicted self-response score)
  - Sparse Spaces (5 percent of the population, 49 percent predicted self-response score)
  - Multicultural Mosaic (14 percent of the population, 45 percent predicted self-response score)
  - Rural Delta and Urban Enclaves (7 percent of the population, 43 percent predicted self-response score)

The CRAT team combined the mindset segmentation and tract segmentation results to produce a detailed summary of how predicted self-response, demographic characteristics, and mindsets toward the census are distributed across the United States. It was found that, first and foremost, no single geographic area or group of people is guaranteed to participate in the 2020 Census. The highest intent to self-respond among the mindsets is just above 80 percent, with 82 percent of people in the Eager Engagers mindset indicating that they intend to participate in the 2020 Census. The highest predicted self-response score among the tract segments is just over 70 percent, with the Responsive Suburbia segment having a 71 percent predicted self-response score. These rates of predicted self-response are above the predicted national average of 60.5 percent and well above the predicted self-response observed for other tract

[28] 'Intend to respond' is based on the 2020 CBAMS Survey question, "If the census were held today, how likely would you be to fill out the census form? (a) Extremely likely, (b) Very likely, (c) Somewhat likely, (d) Not too likely, (e) Not at all likely." The percentages reported above reflect answers of "extremely likely" or "very likely."

segments, such as the Disconnected Doubters, 51 percent of whom intend to respond, and people living in Rural Delta and Urban Enclaves census tracts (43 percent).

The findings in this report underscore the importance of a multifaceted and dynamic communications campaign. The campaign will need to use diverse mass messaging to remind the parts of the country that are more inclined to respond of the benefits of participating in the census, such as determining community funding, and also speak directly to those who are unlikely to self-respond to the 2020 Census due to wariness about the government, the feeling that it does not matter if they participate, or a general lack of knowledge and understanding about the census. Armed with the information described in this and other, related reports, the ICC is working to create messages, craft effective creative materials to convey those messages, determine where and when to allocate advertising, and identify the most efficient division of scarce resources across communications channels and approaches.

9. Acknowledgments

Contributions from Team Y&R were performed under the Integrated Communications Contract YA1323-16-CQ-0003/Order 10 for the 2020 Census awarded to Young & Rubicam. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau or Young & Rubicam.

The results published in this paper have undergone disclosure avoidance review and were released with approvals CBDRB-FY-382 (approved July 16, 2018), CBDRB-FY18-422 (approved August 13, 2018), CBDRB-FY19-198 (March 4, 2019), and CBDRB-FY18-311 (May 23, 2018).

The census staff listed below assisted with their subject matter expertise or overall management of the contract.

Project Management Office
Kendall Johnson
Lizannette Velez
Michelle Green
James Cole
David Earles
Puja Ohri

Decennial Statistical Studies Division
Sarah Konya
Julia Coombs
Kelly Mathews
Michael Bentley
Earl Letourneau
Matt Herbstritt

Decennial Communications and Coordination Office
Kaile Bower

Center for Statistical Research and Methodology
Mary Mulry

Public Information Office
Julie Iriondo

The Team Y&R agencies listed below assisted with various phases of the 2020 Census Predictive Models and Audience Segmentation, this report, and contract management.

PSB
VMLY&R

Reingold
Carol H Williams
Culture ONE World
G+G
The Kālaimoku Group
TDW+Co.
VMLY&R San Juan
Wavemaker

10. References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211.

Andridge, R., & Little, R. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, *78*(1), 40-64.

Baumgardner, S. K., Griffin, D. H., & Raglin, D. A. (2014). *The effects of adding an internet response option to the American Community Survey.* Suitland, MD: U.S. Census Bureau.

Bruce, A., Robinson, J.G., & Sanders, M.V. (2001). Hard-to-Count (HTC) scores and broad demographic groups associated with patterns of response rates in Census 2000. *Annual Meeting of the American Statistical Association*. Atlanta, GA.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211-407.

Erdman, C., & Bates, N. (2017). The low response score (LRS): A metric to locate, predict, and manage hard-to-survey populations. *Public Opinion Quarterly*, *81*(1), 144–156.

Evans, S., Levy, J., Miller-Gonzalez, J., Vines, M., Girón, A. S., Walejko, G., Bates, N., & Trejo, Y. G. (2019). 2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Focus Group final report. Suitland, MD: U.S. Census Bureau. https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2020-report-cbams-focus-group.html

Idrovo, A. J. (2011). Three criteria for ecological fallacy. *Environmental Health Perspectives, 119*(8), a332.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika, 30*(1–2), 81–89.

Letourneau, E. (2018). *Study plan for the 2018 end-to-end census test response rates operational assessment.* Suitland, MD: U.S. Census Bureau.

McGeeney, K., Kriz, B., Mullenax, S., Kail, L., Walejko, G., Vines, M., Bates, N., & Trejo, Y. G. (2019). *2020 Census Barriers, Attitudes, and Motivators Study Survey report*. Suitland, MD: U.S. Census Bureau. https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2020-report-cbams-study-survey.html

Mulry, M., Bates, N., & Virgile, M. (2018). *Viewing participation in censuses and surveys through the lens of lifestyle segments.* Unpublished manuscript, Center for Statistical Research and Methodology, U.S. Census Bureau.

Nielsen. (2019). Local television market universe estimates. Chicago, IL: Author. Retrieved from https://www.nielsen.com/content/dam/corporate/us/en/public%20factsheets/tv/2018-19-dma-ranker.pdf

Phelan, J. (2016). *2015 National Content Test optimizing self-response report.* Suitland, MD: U.S. Census Bureau.

Reddy, C. K., & Vinzamuri, B. (2014). A survey of partitional and hierarchical clustering algorithms. In C. C. Aggarwal & C. K. Reddy (Eds.), *Data clustering algorithms and applications* (pp. 87-110). New York, NY: CRC Press.

Saunders, B. C., Jr. (2012). *2010 Census Integrated Communications Program Mail Response Rates/Take 10 Assessment Report* (2010 Census Planning Memoranda Series). Retrieved from https://www2.census.gov/programs-surveys/decennial/2010/program-management/5-review/cpex/2010-memo-222.pdf

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1-34.

Torrieri, N. (2014). *American Community Survey design and methodology (January 2014)*. Suitland, MD: U.S. Census Bureau. Retrieved from https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch07_2014.pdf

U.S. Census Bureau. (2017). *2020 Census life-cycle cost estimate executive summary.* Retrieved from https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/cost-estimate.html

U.S. Census Bureau. (2018). *2020 Census Detailed Operational Plan for: 12. Internet Self-Response (ISR) Operation.* Retrieved from: https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/ISR-detailed-op-plan.html

U.S. Census Bureau. (2019). *2020 type of enumeration areas (TEAs) – Presentation for IPC*. Suitland, MD: Author.

Valliant, R., Dever, J., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York, NY: Springer.

Warton, D., & Hui, F. (2011). The arcsine is asinine: The analysis of proportions in ecology. *Ecology, 92*(1), 3–10.

Appendix A: Acronyms and Abbreviations

ACS	American Community Survey
ADCOM	Associate Directorate for Communications
AIAN	American Indian and Alaska Native
CAPI	Computer-Assisted Personal Interviewing
CATI	Computer-Assisted Telephone Interviewing
CBAMS	Census Barriers, Attitudes, and Motivators Study
CQA	Census Questionnaire Assistance
CRAT	Communications Research and Analytics Team
CSRM	Center for Statistical Research and Methodology
DMA	Designated Market Area
DRB	Disclosure Review Board
HTC	Hard-to-Count
ICC	Integrated Communications Campaign
ICF	Inner City Fund
IPC	Integrated Partnership and Communications
IPSR	Internet Proportion of Self-Response
IPSRS	Internet Predicted Self-Response Score
ISR	Internet Self-Response
LASSO	Least Absolute Shrinkage and Selection Operator
LRS	Low Response Score
MRI	Mediamark Research and Intelligence
MRR	Mail Response Rates
NCT	National Content Test
NH	Non-Hispanic
NRFU	Nonresponse Followup
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PDB	Planning Database
PSRS	Predicted Self-Response Score
PUMS	Public Use Microdata Sample
TEA	Type of Enumeration Area
WMAE	Weighted Mean Absolute Error

Appendix B: Technical Appendix

B.1 ACS Modeling

B.1.1 Dependent Variables

A government shutdown affected the October 2013 ACS panel and, therefore, was excluded from our calculation. In addition, the CRAT team did not calculate self-check-in rates for tracts in Puerto Rico because the ACS data collection in Puerto Rico is substantially different from most of the United States. Finally, some participants in November and December ACS panels were placed in new census tracts during the following year and therefore had two different tract identifiers. The CRAT team treated these individuals as living in their most recent tract when calculating the ACS overall and internet self-check-in rates.

The U.S. Census Bureau already computes check-in rates by mode for other research, so, following the code and business logic the Census Bureau provided, the CRAT team computed an adapted version of the self-check-in rate for this study in alignment with ACS standard procedures. For the purposes of this study, the CRAT team focused on unassisted self-response to the ACS and did not include telephone questionnaire assistance responses, in-person responses captured through computer-assisted personal interviewing, or computer-assisted telephone interviewing (CATI) returns as types of self-response. The U.S. Census Bureau defines "overall self-check-in rate" as the total number of ACS internet returns before the start of CATI and the number of mail returns up to seven days after the start of CATI[29] from mailable[30] addresses, divided by the total number of mailable housing unit addresses sampled in the tract (excluding group quarters). Tracts that had no mailable addresses in the ACS sample from 2013 to 2016 will have no ACS-PSRS and, therefore, will not have a 2020-PSRS.

Because the CRAT team used raw tract-level self-check-in rates—which are bounded between zero and 100 percent—as our dependent variable, and since linear models can predict self-check-in rates outside the unit interval, it was necessary to apply a logit transformation to raw ACS self-check-in rates that transformed model predictions back to the unit interval.[31] The logit transformation maps proportions to values between negative and positive infinity. Since the logit function is concave for proportions less than 0.5 and convex for those greater than 0.5, small differences in self-check-in rates on either end of the bound result in larger differences on

---

[29] The additional seven days for mail returns gives time for the receipt of responses mailed before the start of CATI.

[30] "The requirement for a 'mailable' address in the United States is met if there is either a complete city-style or rural route address. A complete city-style address includes a house number, street name, and ZIP Code. (The town or city and state fields are not required because they can be derived from the ZIP Code.) A complete rural-route address includes a rural-route number, box number, and ZIP Code. About 97 percent of the 2012 sample addresses in the United States met these criteria and were designated as mailable." (Torrieri, 2014)

[31] The boundedness of the self-check-in rates also violates the assumption of homoscedasticity made by linear approaches. Any rates near either bound necessarily have lower variances than those in the middle. Fitting a linear model on heteroscedastic data yields biased standard errors, which in our case, allows for misinterpretation of variables as significantly affecting self-check-in rates when the actual effect is negligible.

the logit scale. This has the desired effect of producing a more homoscedastic self-response variable for linear modeling. After deriving tract-level predictions on the logit scale, applying the logistic transformation—the inverse of logit—maps the values back to the unit interval, which enables probabilistic reasoning.

One complication with using the logit transformation is that it transforms zero into negative infinity and 1 into positive infinity. To scale back these extreme values, the formula in Equation 3, below, was used. This is the approach described by Warton and Hui (2011), which sets a factor of ε using the values included in x and adds ε to the traditional logit transformation.[32]

$$\text{Logit Transformation} = \log\left(\frac{x+\varepsilon}{1-x+\varepsilon}\right) \qquad (10)$$

B.1.2 Independent Variables

For both the core and expanded variables, the CRAT team used imputation to replace missing values. Several of the expanded variables derived from ACS measures of housing value have a large number of missing values—including 200 to 2,000 tracts with missing values. The median value of tracts in the same county was assigned to any missing value.

B.1.2.1 Core Variables

The core variables are composed of demographic variables identified in the original 2014 LRS study and a set of variables identified as predictive of ACS self-check-in rates by Census Bureau researchers. The CRAT team transformed variables that Erdman and Bates transformed in the original 2014 LRS study using the same logit, square root, or logarithmic transformation functions used in the original study. Table 9 provides a detailed list of all PDB variables included in the list of core variables. Table 10 outlines the non-PDB variables included in core variables.

Table 9. PDB Variables Included in the Core Variables

PDB Variable	Justification	Definition	Transformation
pct_renter_occp_hu_cen	2014 LRS	The percentage of 2010 Census occupied housing units that are not owner occupied, whether they are rented or occupied without payment of rent	Logit
pct_pop_18_24_cen	2014 LRS	The percentage of the 2010 Census total population between 18 and 24 years of age	Square Root
pct_female_no_hb_cen	2014 LRS	The percentage of all ACS occupied housing units with a	Square Root

[32] Warton and Hui (2011) suggest "taking ε as the minimum non-zero proportion y, or if proportions are large, the minimum non-zero proportion for 1 – y." The CRAT team used the smaller of these two suggested adjustment factors.

		female householder and no husband of householder present	
pct_nh_white_alone_cen	2014 LRS	The percentage of the ACS population that indicates no Hispanic origin and their only race as "White" or reports entries such as Irish, German, Italian, Lebanese, Arab, Moroccan, or Caucasian	Logit
pct_pop_65_plus_cen	2014 LRS	The percentage of the 2010 Census total population that is 65 years or older	Square Root
pct_rel_under_6_cen	2014 LRS	The percentage of 2010 Census family-occupied housing units with a related child under six years of age; same-sex couple households with no relatives of the householder present are not included in the denominator	Square Root
pct_males_cen	2014 LRS	The percentage of the 2010 Census total population that is male	None
pct_mrdcple_hhd_cen	2014 LRS	The percentage of all 2010 Census occupied housing units where the householder and his or her spouse are listed as members of the same household; does not include same-sex married couples	None
pct_pop_25_44_cen	2014 LRS	The percentage of the ACS population that is between 25 and 44 years of age	Square Root
pct_vacant_units_cen	2014 LRS	The percentage of all 2010 Census housing units that have no regular occupants on Census Day; housing units with their usual occupants temporarily away (such as on vacation, a business trip, or in the hospital) are not considered vacant, but	Logarithm

		housing units temporarily occupied on Census Day by people who have a usual residence elsewhere are considered vacant	
pct_college_acs	2014 LRS	The percentage of the ACS population 25 years or older that have a college degree or higher	Logit
med_hhd_inc_acs	2014 LRS	Median ACS household income for the tract	Logarithm
pct_pop_45_64_cen	2014 LRS	The percentage of the 2010 Census total population that is between 45 and 64 years of age	Square Root
avg_tot_prns_in_hhd_cen	2014 LRS	The average number of persons per ACS occupied housing unit; this was calculated by dividing the total household population in the ACS by the total number of occupied housing units in the ACS	Logarithm
pct_hhd_moved_in_acs	2014 LRS	The percentage of all ACS occupied housing units where the householder moved into the current unit in the year 2010 or later	Square Root
pct_hispanic_cen	2014 LRS	The percentage of the 2010 Census total population that identifies as "Mexican," "Puerto Rican," "Cuban," or "another Hispanic, Latino, or Spanish origin"	Logit
pct_single_unit_acs	2014 LRS	The percentage of all ACS housing units in a structure that contains only that single unit	Logit

popdensity_cen[33]	2014 LRS	The total population in the 2010 Census divided by land area in square miles	Logarithm
pct_prs_blw_pov_level_acs	2014 LRS	The percentage of the ACS eligible population classified as below the poverty level given total family or household income within the last year, family size, and family composition	Square Root
pct_diff_hu_1yr_ago_acs	2014 LRS	The percentage of the ACS population aged one year or older that moved from another residence in the U.S. or Puerto Rico within the last year	Square Root
pct_pop_5_17_cen	2014 LRS	The percentage of the 2010 Census total population that is between five and 17 years of age	Square Root
pct_nh_blk_alone_cen	2014 LRS	The percentage of the 2010 Census total population that indicates no Hispanic origin and their only race as "Black, African Am., or Negro" or report entries such as African American, Kenyan, Nigerian, or Haitian	Logit
pct_sngl_prns_hhd_cen	2014 LRS	The percentage of all 2010 Census occupied housing units where a householder lives alone	Square Root
pct_not_hs_grad_acs	2014 LRS	The percentage of the ACS population 25 years or older that is not a high school graduate and has not received a diploma or the equivalent	Square Root
med_house_value_acs	2014 LRS	Median of ACS respondents' house value estimates for the tract	Logarithm

[33] This variable was calculated for this analysis using two other variables in the Planning Database (Tot_Population_CEN_2010 and LAND_AREA).

pct_othr_lang_acs	CSRM	The percentage of the ACS population five years or older that speaks a language other than English at home	Square Root
pct_crowd_occp_u_acs	CSRM	The percentage of ACS occupied housing units that have more than 1.01 persons per room	Square Root
pct_no_ph_srvc_acs	CSRM	The percentage of ACS occupied housing units that do not have a working telephone and available service	Square Root
pct_mobile_homes_acs	CSRM	The percentage of all ACS housing units considered mobile homes	None

Source: Census Planning Database. https://www.census.gov/research/data/planning_database/

Table 10. Non-PDB Variables Included in the Core Variables

Variable	Justification	Definition	Transformation
Esri	(Mulry, Bates, and Virgile, 2018)	The Esri Tapestry Segmentation is a market segmentation database that groups neighborhoods with similar psychographic, social, and demographic characteristics into one of 68 distinct segments.	None
nonmobile	CSRM	The percentage of the ACS population that indicated it had access to the internet through a nonmobile device	None
access	CSRM	The percentage of the ACS population that indicated it had access to the internet through any means	None
urban	CSRM	The proportion of land area in urban blocks	None

Source: Esri Tapestry Segmentation https://www.esri.com/en-us/arcgis/products/tapestry-segmentation/overview, American Community Survey. https://www.census.gov/programs-surveys/acs/

Table 11 details the ACS questions and response options used to create the internet access measures.

Table 11. ACS Internet Access and Means of Internet Access Questions

	Years	Question	Answer Choices
ACS Internet Access Questions	2013–2015	Q10. At this house, apartment, or mobile home – do you or any member of this household access the Internet?	a) Yes, with a subscription to an Internet service (checkbox) b) Yes, without a subscription to an Internet service (checkbox, if yes skip to Q12) c) No Internet access at this house, apartment, or mobile home (checkbox, if yes skip to Q12)
	2016–2018	Q9. At this house, apartment, or mobile home – do you or any member of this household have access to the Internet?	a) Yes, by paying a cell phone company or Internet service provider (checkbox) b) Yes, without paying a cell phone company or Internet service provider (checkbox, skip to Q11) c) No access to the Internet at this house, apartment, or mobile home (checkbox, skip to Q11)
ACS Means of Internet Access Questions	2013–2015	Q11. At this house, apartment, or mobile home – do you or any member of this household subscribe to the Internet using…	a) Dial-up service? (yes/no) b) DSL service? (yes/no) c) Cable modem service? (yes/no) d) Fiber-optic service? (yes/no) e) Mobile broadband plan for a computer or a cell phone? (yes/no) f) Satellite Internet service? (yes/no) g) Some other service? (yes/no, specify service)
	2016–2018	Q10. Do you or any member of this household have access to the Internet using a…	a) Cellular data plan for a smartphone or other mobile device (yes/no) b) Broadband (high speed) Internet service such as cable, fiber optic, or DSL service installed in this household? (yes/no) c) Satellite Internet service installed in this household? (yes/no) d) Dial-up Internet service installed in this household? (yes/no) e) Some other service? (yes/no, specify service)

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/

B.1.2.2 Expanded Variables

The expanded variables included all core variables, as well as the entire set of tract-level PDB variables[34] and all appropriate transformations of each variable. These transformations included square root and logarithm transformations for all variables and logit transformations for all proportion variables. The CRAT team slightly adjusted the standard logarithm transformation to account for the fact that the logarithm of zero is undefined and added γ (the smallest nonzero value of the variable) before taking the logarithm. This ensured that the transformation was defined for all values. In applying the logit transformation, the CRAT team used the same approach as for the self-check-in rate. The CRAT team did not alter the standard square root transformation.

$$\text{Logarithm Transformation} = \log(x + \gamma) \qquad (11)$$

B.1.3 Modeling Algorithms

The CRAT team estimated four types of models to determine the best way to predict overall and internet ACS self-check-in rates. The sections below describe three of these models.

B.1.3.1 Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO model combines the linear estimation of the ordinary least squares (OLS) regression with shrinkage parameters that reduced the impact of irrelevant predictors. More specifically, the LASSO model used was the L1-penalized regression model. The penalization aspect of this model selects a limited set of predictive variables from the full set of input variables and puts all other coefficients at zero. The CRAT team used k-fold cross-validation to identify alpha, the penalty parameter of the model that minimizes the training error. This allowed the LASSO model to include a large number of independent variables, from which the model removed irrelevant ones.

B.1.3.2 Random Forest

The random forest algorithm is an ensemble model built from a collection of decision trees, each fit to a different subset of the data. Each tree "votes" for the predicted outcome. The model can identify complex interactions and does not require prescaling of variables.

B.1.3.3 Gradient-Boosted Trees

The gradient-boosted trees algorithm is an ensemble model built from a series of shallow decision trees. Each subsequent tree improves accuracy by correcting the predictions of the previous tree. The model can identify complex interactions and does not require prescaling of variables.

---

[34] The CRAT team produced models using both the 2016 and 2014 PDBs, as described in the validation section, so only those variables present in both vintages of the PDB were included as part of the expanded set of variables. The CRAT team also removed PDB variables that captured the amount of self-response to the 2010 Census, such as the LRS or the 2010 mail return rate, because direct measures of self-response are inappropriate in models that predict self-response.

B.1.4 Model Validation

Because the demographic variables measured by the PDB shift over time, it was best to use the most recent vintage of the PDB available. However, each PDB uses five years of ACS responses to create its variables, which potentially lets years of ACS data included in a PDB overlap with or come after the years of ACS data used to create self-check-in rates. To avoid this, the CRAT team used separate vintages of the PDB for training and validation. This ensured that each process used the latest available PDB vintage that did not overlap with the years of ACS data used to calculate self-check-in rates. Table 12 describes the specific composition of the training and validation data sets.

Table 12. Temporal Distribution of Training and Validation Data Sets

Data Set	Years of ACS Data Included in Self-Check-In Rates	Years of ACS Data Included in PDB Variables
Initial Training Data Set	2013–2015	2008–2012
Validation Data Set	2016	2010–2014

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/,
Census Planning Database. https://www.census.gov/research/data/planning_database/

The primary metric for model validation was the weighted mean absolute error (WMAE) between the actual 2016 ACS self-check-in rate and the predicted 2016 ACS self-check-in rate. Calculating WMAE using the number of households in the tract as the weight ensured that accuracy in tracts with very small numbers of households did not overly influence model selection. Using WMAE, as opposed to the weighted mean squared error metric, made model selection less susceptible to large errors in a small number of tracts. These choices are particularly important because validating models with a single year of ACS data increased the potential for tracts with small numbers of cases that could cause outliers, particularly in tracts with fewer households.

While the WMAE from the validation data set was the primary determinant for model selection, the CRAT team also validated models for other types of error using the training error and rank-ordering error.[35] Once the WMAE, training error, and Kendall's tau for each model was calculated, all three metrics were used to determine which approach would be best for producing the final ACS Predicted Self-Response Score (PSRS) and ACS Internet Predicted Self-Response Score (IPSRS).

B.1.5 ACS Model Selection

Table 13 shows error metrics for all four model types for both overall and internet self-check-in rates.

Table 13. Error Metrics for Candidate ACS-PSRS Models

[35] To calculate the training error, the CRAT team computed the WMAE based on k-fold cross-validation of the training data. This measure tells us how well the model fits the training data set. To calculate the rank-ordering error, Kendall's tau was used as a measure of rank correlation between the actual and predicted self-check-in rates (Kendall, 1938; see also Idrovo, 2011). This measure was useful for comparing the predicted and actual orders of tracts in terms of self-check-in rate.

	Candidate Model	Validation Error	Training Error	Kendall's Tau
Overall	Least Absolute Shrinkage and Selection Operator (LASSO)	0.0625	0.0462	0.6719
	Random Forest	0.0631	0.0432	0.6676
	Gradient Boosted Trees	0.0652	0.0459	0.6583
	Ordinary Least Squares (OLS)	0.0700	0.0503	0.6313
Internet	Least Absolute Shrinkage and Selection Operator (LASSO)	0.0558	0.0358	0.6621
	Random Forest	0.0559	0.0369	0.6668
	Gradient Boosted Trees	0.0563	0.0348	0.6650
	Ordinary Least Squares (OLS)	0.0590	0.0373	0.6475

Approved for release by the Disclosure Review Board on March 4, 2019, with approval number CBDRB-FY19-198

B.1.6 Disclosure Review Board Process

The CRAT team evaluated model fit metrics on the final model for a range of privacy parameter values and requested a privacy parameter from the Census Bureau's Disclosure Review Board that produced accurate results while also protecting the Title 13 data. Following the approval of our selected privacy parameter, the CRAT team re-estimated the ACS-PSRS and ACS-IPSRS models using noise-injected dependent variables and ACS internet variables. The results of these noise-injected models were the final ACS-PSRS, ACS-IPSRS, and ACS-IPSR modeled scores. These results, their noise-injected inputs, and any further products developed from them can all be made publicly available without risking disclosure of Title 13 data.[36]

B.1.7 Scoring and Final Estimation

Because of the larger range of years of ACS data included in the scoring data set, the appropriate earlier vintage of the PDB was used to re-estimate these models (see Table 14). Tracts that were included in the 2013 through 2016 ACS sample but that did not have mailable addresses did not receive scores.

Table 14. Temporal Distribution of Scoring Data Set

Data Set	Years of ACS Data Included in Self-Check-In Rates	Years of ACS Data Included in PDB Variables
Final Scoring Data Set	2013–2016	2008–2012

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/,
Census Planning Database. https://www.census.gov/research/data/planning_database/

After creating the ACS-PSRS and ACS-IPSRS, the ratio of these two scores was used to create the ACS-IPSR (see following equation). The ACS-IPSR indicates the predicted proportion of a tract's

[36] In addition to approving the privacy parameter and noise-injection procedure, the Census Bureau's Disclosure Review Board (DRB) approved the release of two national-level aggregates from protected 2015 National Content Test and 2017 Census Test data for the 2020-IPSR adjustment. This release received DRB approval on May 23, 2018, with approval number CBDRB-FY18-311.

overall self-check-in to the ACS that occurs online based on the tract's characteristics. Recall that the ACS-IPSR was an intermediary research product intended to be adjusted to create the 2020-IPSR, which will be used to inform the communications campaign about the proportion of a tract's self-response to the 2020 Census that is likely to happen online.

$$\text{ACS-IPSR}_{tract} = \frac{\text{ACS-IPSRS}_{tract}}{\text{ACS-PSRS}_{tract}} \qquad (12)$$

B.1.8 ACS Model Results

Figure 37 shows the distribution of predicted unweighted ACS-PSRS scores. The distribution lacks symmetry and it is approaching, but not reaching, a bimodal distribution with a smaller set of tracts amassing near the 12 percent to 25 percent region and a larger set of tracts hovering in the 37 percent to 50 percent region. The unweighted mean and variance ACS-PSRS score are 39 and 190, respectively.

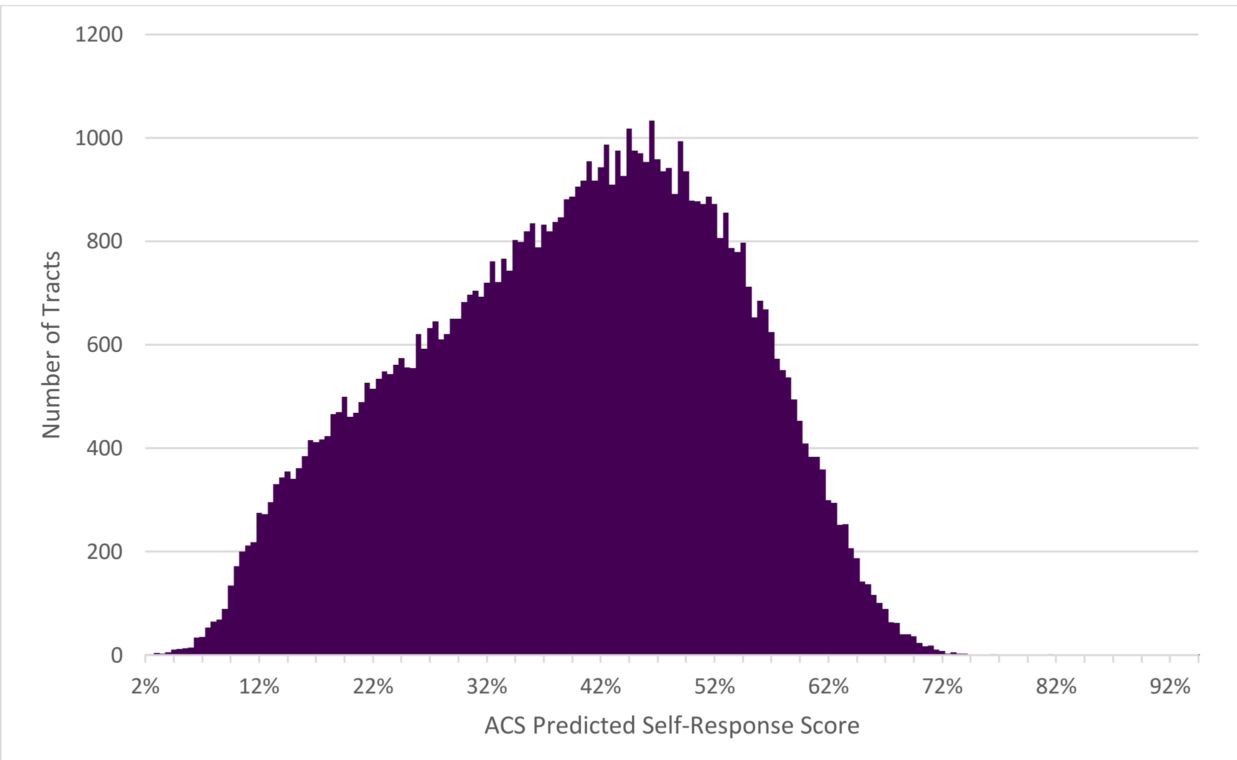


Figure 37. Unweighted ACS-PSRS Distribution

Figure 38 shows a weighted version of the distribution. The CRAT team constructed weights for each tract using the ACS estimated total occupied housing units by:

$$PSRSwt_i = \frac{Total\ Occupied\ Units_i}{\sum Total\ Occupied\ Units_i}$$

The weighted mean and variance for the ACS-PSRS are 41 and 169 respectively. The CRAT team used total occupied housing units as a measure of households and chose this as a weight because response is a household-level characteristic. The weighted distribution gives greater importance to tracts with more households when evaluating the overall distribution and does not result in a dramatic change in the shape of the distribution.

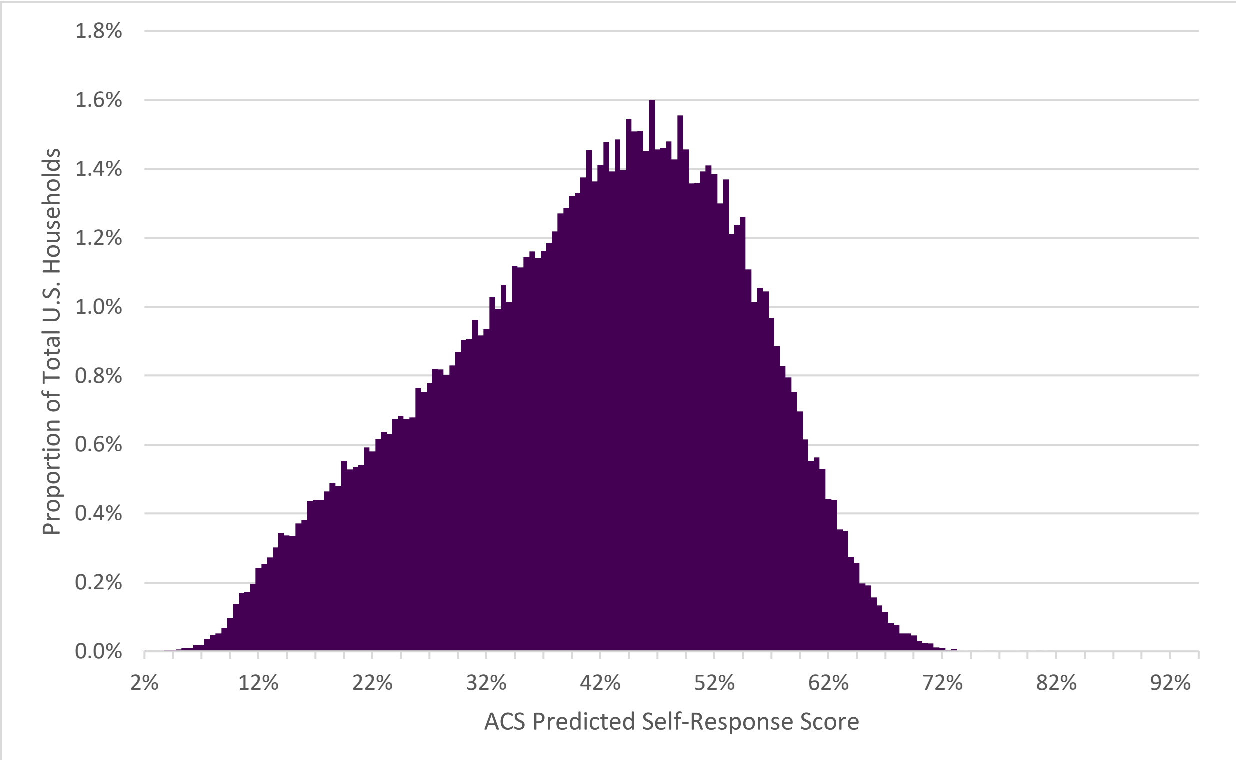


Figure 38. Weighted ACS-PSRS Distribution

B.2 Adjusting the ACS-IPSR

To determine the correct size for adjustments, the proportion of internet self-response for tracts in the 2015 National Content Test[37]and 2017 Census Test was calculated. When conducting these tests, the Census Bureau did not use a sampling design optimized to create tract-level estimates; therefore, the sample allocation and number of respondents per tract vary widely. Tracts with small numbers of respondents had high sampling error for their self-response estimates. As a result, the CRAT team excluded tracts that did not meet a minimum number of respondents based on a predetermined margin of error for a hypothetical proportion that does not include the originating surveys' sample design or weights.[38] Excluding

[37] The CRAT team used only the panels of the 2015 National Content Test that were most similar to the Internet First and Internet Choice contact strategies. The 2015 NCT included six panels that were slight variations on the Internet First contact strategy, but two had more significant deviations from the Internet First contact strategy and were excluded from this adjustment. The CRAT team used only Panel 5, which is explicitly identified as Internet Choice, for the Internet Choice adjustment.

[38] The relationship between sample size and margin of error for a proportion depends on the value of the proportion and the confidence level chosen. A proportion of 50 percent gives the largest possible variance and, therefore, the most conservative possible sample size estimate. This makes the relationship between sample size and margin of error, confidence level, and estimated proportion n=z_(1-α/2)^2×p(1-p)/e^2 (Valliant, 2013). For a 90 percent confidence level, the z-score is 1.645, and a sample size of 68 exceeds a 10 percent margin of error threshold.

all tracts with 68 or fewer respondents removes those whose maximum margin of error is larger than 10 percent with a 90 percent confidence level, for a hypothetical proportion equaling 50 percent.

The separate adjustments by contact strategy, which lower the predicted proportion of online self-response for households in tracts that receive a questionnaire in their initial mailing (Internet Choice) and raise the predicted proportion of online self-response for households that do not (Internet First), dramatically alters the ACS-based scores. Figure 39 illustrates that these adjustments result in two distinct sets of scores, each with its own local maximum. Whereas most tracts have ACS-IPSR values between 50 and 70 percent, few tracts have 2020-IPSR values in that range. Moreover, the two distinct sets of modeled 2020-IPSR scores each have smaller variances than the combined set of ACS-IPSR scores.

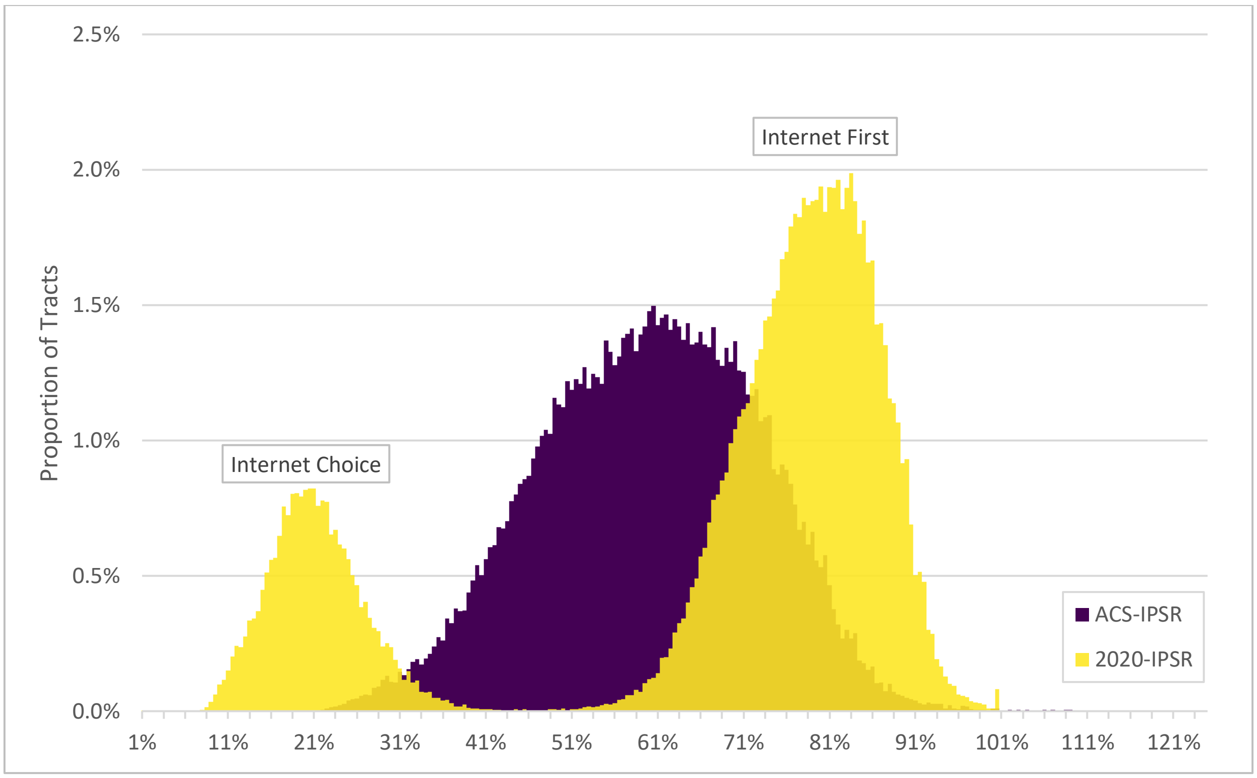


Figure 39. Comparison of the ACS-IPSR and 2020-IPSR

B.3 Creating Tract-Level Self-Response Rates for the 2010 Census

To understand the potential variation in tract-level timing of 2020 Census self-response, the CRAT team adjusted tract-level participation rates reported during the 2010 Census into self-response rates. The Take 10 Program posted preliminary participation rates on the 2010 Census website from March 22, 2010 through April 28, 2010. This period started after the Census Bureau began internal reporting via the mail response rates (MRR) program on March 6, 2010, and it ended before the start of Nonresponse Followup (NRFU) operations on May 1, 2010. The Census Bureau updated these rates daily on weekdays during this period. The rates reported were preliminary participation rates, which the Census Bureau calculated as the percentage of the valid mail-back universe that completed questionnaires (the valid mail-back universe excludes all undeliverable-as-addressed addresses).

The tract identifiers used to report participation rates during the 2010 Census campaign reflected the geographic definitions used at the time, which were based on the 2000 Census. To convert participation rates into self-response rates and have appropriate weights to use in the assessment of those rates, the CRAT team converted these tract identifiers into the tract definitions that the Census Bureau introduced following the 2010 Census.[39]

The CRAT team updated the tract identifiers using a crosswalk provided by the Census Bureau.[40] For each intersection between old and new tracts, this file contains fields indicating the tract identifier from 2000, the tract identifier from 2010, and a variable indicating the percentage of the housing units in 2010 that were included in the intersection between the two geographies. Because the reported participation rates are for entire tracts, the CRAT team combined portions of different tracts by assuming that each part of a tract shared the same rate as the tract as a whole. Finally, the CRAT team transformed participation rates into self-response rates by multiplying each participation rate by the number of valid mail-back addresses in that tract and then dividing by the total number of mail-back addresses in the tract. Because the Census Bureau reported participation rates prior to the beginning of NRFU operations, these are self-response rates.

B.4 Mindset Data Preparation

The CRAT team produced mindsets as described in Section 5. The following variables found in Table 15 were inputs to a PCA, which underwent a varimax rotation to produce the factor loadings in Table 16. The factors listed in Table 15 indicate that the variable had a factor loading of at least 0.4 on that factor after performing the varimax rotation. Three variables listed at the bottom of Table 15 did not have any factor loadings of at least 0.4. Additionally, Table 17 lists eight inputs that the CRAT team removed prior to running the final PCA based on preliminary PCA analyses showing that these variables do not strongly load on any components.

Table 15. Description of Mindset Variables

2020 CBAMS Variable Name	2020 CBAMS Question [response option for dummy variable]	Component
civic_participation_1	Which of the following have you ever done, if any? [Voted in an election]	1
civic_participation_2	Which of the following have you ever done, if any? [Signed a petition]	1
civic_participation_3	Which of the following have you ever done, if any? [Posted your own thoughts or comments on political or social issues online]	1

[39] In some cases, the Census Bureau updated tract identifiers following the 2010 Census. The CRAT team accounted for these changes when updating tract identifiers to match the most recent vintage of the PDB.
[40] https://www.census.gov/geo/maps-data/data/tract_rel_layout.html

civic_participation_4	Which of the following have you ever done, if any? [Volunteered at any organization]	1
civic_participation_5	Which of the following have you ever done, if any? [Worn a button/ bracelet/pin for an issue or cause]	1
civic_participation_6	Which of the following have you ever done, if any? [Contacted, or attempted to contact, a politician or civil servant to express your views]	1
civic_participation_7	Which of the following have you ever done, if any? [Attended a neighborhood or community meeting]	1
civic_participation_8	Which of the following have you ever done, if any? [Participated in an organized protest or rally of any kind]	1
civic_participation_9	Which of the following have you ever done, if any? [Donated money or raised funds for social or political activity]	1
a_motivator1	How important, if at all, is each of the following programs and services to you personally? [Daycare for children]	2
a_motivator5	How important, if at all, is each of the following programs and services to you personally? [Job training programs]	2
a_motivator7	How important, if at all, is each of the following programs and services to you personally? [Public transportation]	2
a_motivator8	How important, if at all, is each of the following programs and services to you personally? [Schools and the education system]	2
p_motivator1	How important, if at all, is each of the following to you personally? [Showing you are proud of your cultural heritage]	2
p_motivator2	How important, if at all, is each of the following to you personally? [Contributing to a better future for your community]	2
p_motivator4	How important, if at all, is each of the following to you personally? [That civil rights laws are enforced]	2
use10	How important, if at all, is each of these uses to you personally? [Providing information for your local government to plan for changes in your community]	2
familiarity	How familiar are you with the U.S. census?	3
participation1	If the census were held today, how likely would you be to fill out the census form?	3
participation2	Thinking about most people you know, if the census were held today how likely would they be to fill out the census form?	3
participation4	How likely are you to encourage someone you know to fill out the 2020 Census form?	3
use9	How important, if at all, is each of these uses to you personally? [Determining how many elected representatives your state has in Congress]	3, 8

efficacy	How much, if at all, do you think it matters if you personally are counted in the 2020 Census?	3
a_motivator2	How important, if at all, is each of the following programs and services to you personally? [Fire departments]	4
a_motivator3	How important, if at all, is each of the following programs and services to you personally? [Police departments]	4
a_motivator4	How important, if at all, is each of the following programs and services to you personally? [Hospitals and healthcare]	4
a_motivator6	How important, if at all, is each of the following programs and services to you personally? [Roads and highways]	4
p_motivator3	How important, if at all, is each of the following to you personally? [Fulfilling your civic duty]	4
internet1	About how often do you use the internet?	5
internet2_2 internet2_3	Which devices do you often use to access the internet? [Smartphone OR tablet computer only]	5
Internet2_1 internet2_2 internet2_3	Which devices do you often use to access the internet? [Desktop or laptop computer OR smartphone OR tablet computer]	5
internet3_1	Which of the following comes closest to your view? [I prefer to fill out paper forms instead of online forms]	5
internet3_2	Which of the following comes closest to your view? [I prefer to fill out online forms instead of paper forms]	5
trust_federal	How much of the time do you think you can trust the FEDERAL government to do what is right?	6
trust_state	How much of the time do you think you can trust your STATE government to do what is right?	6
trust_local	How much of the time do you think you can trust your LOCAL government to do what is right?	6
concern1	How concerned are you, if at all, that the Census Bureau will not keep answers to the 2020 Census confidential?	7
concern2	How concerned are you, if at all, that the Census Bureau will share answers to the 2020 Census with other government agencies?	7
concern3	How concerned are you, if at all, that the answers you provide to the 2020 Census will be used against you?	7
use1_1	Is the census used to decide how much money communities will get from the government, or is it not used for this? [Yes, used for this]	8
use2_1	Is the census used to decide how many representatives each state will have in Congress, or is it not used for this? [Yes, used for this]	8
use4_2	Is the census used to determine property taxes, or is it not used for this? [No, not used]	8

use5_2	Is the census used to help the police and FBI keep track of people who break the law, or is it not used for this? [No, not used]	8
use6_2	Is the census used to locate people living in the country without documentation, or is it not used for this? [No, not used]	8
use8_2	Is the census used to determine the rate of unemployment, or is it not used for this? [No, not used]	8
participation3	Based on your past experience or just your best guess, how long do you think it would take you personally to fill out the 2020 Census?	None
use3_1	Is the census used to see what changes have taken place in the size, location, and characteristics of the people in the United States, or is it not used for this? [Yes, used for this]	None
election2016	Thinking about the 2016 general election for President and other offices, did you happen to vote in the election, or did things come up that kept you from voting?	None

Source: 2020 CBAMS Survey

Table 16. Factor Loadings of Mindset Variables

2020 CBAMS Variable Name	Component 1	2	3	4	5	6	7	8
familiarity	-0.17	0.00	0.63	0.03	-0.04	-0.08	-0.04	-0.04
participation1	-0.15	0.00	0.79	0.08	-0.09	0.05	-0.11	0.02
participation2	-0.01	-0.02	0.70	0.02	-0.01	0.12	-0.04	0.06
participation3	0.02	0.01	0.14	-0.01	-0.17	0.00	-0.11	0.03
participation4	-0.07	0.21	0.71	0.01	-0.05	0.11	-0.04	0.01
internet1	-0.19	0.02	0.12	0.06	-0.84	-0.01	-0.02	-0.04
a_motivator1	0.05	0.70	-0.05	0.08	-0.02	0.03	0.00	-0.06
a_motivator2	-0.01	0.21	0.06	0.83	-0.04	0.02	-0.01	0.00
a_motivator3	0.03	0.14	0.07	0.83	-0.03	0.08	-0.01	0.01
a_motivator4	-0.03	0.26	0.04	0.75	-0.07	0.02	-0.01	-0.02
a_motivator5	0.05	0.77	0.00	0.14	0.04	-0.01	0.05	-0.05
a_motivator6	0.00	0.33	0.13	0.61	-0.09	0.01	0.02	-0.01
a_motivator7	0.01	0.73	0.06	0.08	0.05	0.05	0.05	-0.06
a_motivator8	-0.04	0.66	0.01	0.26	-0.09	0.02	-0.06	-0.03
p_motivator1	0.09	0.53	0.08	0.18	0.14	0.09	0.19	-0.04
p_motivator2	-0.12	0.56	0.22	0.35	-0.08	0.13	0.03	-0.04
p_motivator3	-0.16	0.33	0.38	0.41	0.04	0.16	0.03	0.02

p_motivator4	-0.12	0.48	0.16	0.39	-0.09	0.10	0.02	-0.01
use9	-0.24	0.23	0.48	0.32	0.03	0.12	0.08	-0.07
use10	-0.14	0.41	0.40	0.36	-0.05	0.16	0.04	-0.06
trust_federal	0.12	0.10	0.11	0.05	0.01	0.84	-0.04	-0.02
trust_state	0.05	0.11	0.09	0.06	-0.03	0.89	-0.02	-0.02
trust_local	-0.02	0.07	0.12	0.09	-0.05	0.85	-0.04	-0.02
concern1	0.07	0.05	0.00	0.03	0.09	-0.01	0.88	-0.04
concern2	0.07	0.06	-0.02	0.01	0.09	-0.04	0.89	-0.03
concern3	0.12	0.08	-0.12	-0.02	0.02	-0.05	0.81	-0.05
efficacy	-0.11	0.17	0.55	0.17	-0.01	0.18	0.14	-0.01
internet2_2, internet2_3	0.11	-0.10	0.01	-0.02	0.68	0.01	0.06	-0.01
internet2_1, internet2_2, internet2_3	0.12	0.01	-0.12	-0.04	0.76	0.04	-0.03	0.01
internet3_1	-0.12	-0.05	-0.01	0.04	-0.71	0.05	-0.04	-0.02
internet3_2	0.08	0.05	0.03	-0.05	0.62	-0.06	0.04	0.07
use1_1	0.17	0.09	-0.37	-0.02	0.08	0.05	0.06	0.47
use2_1	0.24	0.22	-0.40	-0.06	0.09	0.03	0.07	0.41
use3_1	0.18	0.13	-0.35	-0.13	0.21	0.01	0.13	0.37
use4_2	-0.03	-0.09	0.03	0.01	0.04	0.00	-0.02	0.63
use5_2	-0.08	-0.09	0.10	0.02	-0.09	-0.06	-0.12	0.60
use6_2	-0.08	-0.10	0.04	0.03	-0.11	-0.01	-0.11	0.57
use8_2	0.07	-0.06	-0.03	-0.04	0.15	-0.01	0.05	0.59
civic_participation_1	0.44	0.24	-0.23	-0.22	-0.08	0.10	0.01	-0.05
civic_participation_2	0.65	0.06	-0.09	-0.07	0.20	0.07	0.06	-0.02
civic_participation_3	0.55	-0.09	0.00	0.07	0.22	0.08	0.04	0.03
civic_participation_4	0.63	0.07	-0.05	-0.09	0.20	-0.01	0.08	0.03
civic_participation_5	0.64	-0.10	0.02	-0.02	0.08	-0.02	0.03	-0.01
civic_participation_6	0.66	0.06	-0.18	0.00	0.03	0.05	0.02	0.01
civic_participation_7	0.61	0.02	-0.16	-0.05	-0.01	-0.04	0.01	0.01
civic_participation_8	0.61	-0.15	-0.06	0.13	0.05	0.00	0.01	0.02
civic_participation_9	0.67	-0.02	-0.08	0.01	0.07	-0.05	0.06	0.03
election2016	0.39	0.20	-0.29	-0.18	-0.10	0.10	-0.03	-0.04

Approved for release by the Disclosure Review Board on March 4, 2019, with approval number CBDRB-FY19-198

Table 17. Variables Eliminated From Mindset Creation

2020 CBAMS Variable Name	2020 CBAMS Question [response option for dummy variable]
use7	Does the census count both citizens and non-citizens, or only citizens?
legal1	Does the law require you to answer the census questions, or is this not required by law?
legal2	Is the Census Bureau required by law to keep information confidential, or is this not required by law?
legal3	Does the U.S. Constitution require that the census be conducted, or is this not something the Constitution requires?
benefit_harm_community_1	Do you believe that answering your 2020 Census form could benefit or harm YOUR COMMUNITY in any way? [Benefit]
benefit_harm_community_2	Do you believe that answering your 2020 Census form could benefit or harm YOUR COMMUNITY in any way? [Harm]
benefit_harm_personal_1	Do you believe that answering your 2020 Census form could personally benefit or harm YOU in any way? [Benefit]
benefit_harm_personal_2	Do you believe that answering your 2020 Census form could personally benefit or harm YOU in any way? [Harm]

Source: 2020 CBAMS Survey

B.5 Tract Segmentation Data Preparation

Tract segmentation variables consist of variables gathered from four data sources. Variables selected from the 2016 PDB are the ones CRAT previously identified as most predictive of tract-level self-response. The 2020-PSRS and 2020-IPSR are predictions of tract-level self-response and the internet proportion of self-response. The ACS Internet Access and Nonmobile Internet Access variables are noise-injected four-year estimates of internet access produced using responses to the 2013-2016 ACS. Lastly, the timing flag is a variable measuring the days taken for 50% of a given tract's population to respond to the 2010 Census produced from the 2010 Census Take 10 data.

Table 18. Description of Tract Segmentation Variables

2020 Segmentation Variable Name	Description	Source
pct_renter_occp_hu_cen__logit	The percentage of 2010 Census occupied housing units that are not owner occupied, whether they are rented or occupied without payment of rent (logit transformed)	2016 PDB Core Variable
pct_pop_18_24_cen__sqrt	The percentage of the 2010 Census total population that is between 18 and 24 years old (square root transformed)	2016 PDB Core Variable
pct_female_no_hb_cen__sqrt	The percentage of all 2010 Census occupied housing units with a female householder and no husband of	2016 PDB Core Variable

	householder present (square root transformed)	
pct_nh_white_alone_cen__logit	The percentage of the 2010 Census total population that indicate no Hispanic origin and their only race as "White" or report entries such as Irish, German, Italian, Lebanese, Arab, Moroccan, or Caucasian (logit transformed)	2016 PDB Core Variable
pct_pop_65plus_cen__sqrt	The percentage of the 2010 Census total population that is 65 years old or over (square root transformed)	2016 PDB Core Variable
pct_rel_under_6_cen__sqrt	The percentage of 2010 Census family-occupied housing units with a related child under 6 years old; same-sex couple households with no relatives of the householder present are not included in the denominator (square root transformed)	2016 PDB Core Variable
pct_males_cen	The percentage of the 2010 Census total population that is male	2016 PDB Core Variable
pct_mrdcple_hhd_cen	The percentage of all 2010 Census occupied housing units where the householder and his or her spouse are listed as members of the same household; does not include same sex married couples	2016 PDB Core Variable
pct_pop_25_44_cen__sqrt	The percentage of the 2010 Census total population that is between 25 and 44 years old (square root transformed)	2016 PDB Core Variable
pct_vacant_units_cen__log	The percentage of all 2010 Census housing units that have no regular occupants on Census Day; housing units with its usual occupants temporarily away (such as on vacation, a business trip, or in the hospital) are not considered vacant, but housing units temporarily occupied on Census Day by people who have a usual residence elsewhere are considered vacant (log transformed)	2016 PDB Core Variable
pct_college_acs__logit	The percentage of the ACS population aged 25 years and over that have a college degree or higher (logit transformed)	2016 PDB Core Variable

med_hhd_inc_acs__log	Median ACS household income for the tract (log transformed)	2016 PDB Core Variable
pct_pop_45_64_cen__sqrt	The percentage of the 2010 Census total population that is between 45 and 64 years old (square root transformed)	2016 PDB Core Variable
avg_tot_prns_in_hhd_cen__log	The average number of persons per 2010 Census occupied housing unit. This was calculated by dividing the total household population in the 2010 Census by the total number of occupied housing units in the 2010 Census. (log transformed)	2016 PDB Core Variable
pct_hhd_moved_in_acs__sqrt	The percentage of all ACS occupied housing units where the householder moved into the current unit in the year 2010 or later (square root transformed)	2016 PDB Core Variable
pct_hispanic_cen__logit	The percentage of the 2010 Census total population that identify as "Mexican", "Puerto Rican", "Cuban", or "another Hispanic, Latino, or Spanish origin" (logit transformed)	2016 PDB Core Variable
pct_single_unit_acs__logit	The percentage of all ACS housing units that are in a structure that contains only that single unit (logit transformed)	2016 PDB Core Variable
popdensity_cen__log	The 2010 Census tract population divided by that tract’s land area in squared miles (log transformed)	2016 PDB Core Variable
pct_prs_blw_pov_lev_acs__sqrt	The percentage of the ACS eligible population that are classified as below the poverty level given their total family or household income within the last year, family size, and family composition (square root transformed)	2016 PDB Core Variable
pct_diff_hu_1yr_ago_acs__sqrt	The percentage of the ACS population aged 1 year and over that moved from another residence in the U.S. or Puerto Rico within the last year (square root transformed)	2016 PDB Core Variable
pct_pop_5_17_cen__sqrt	The percentage of the 2010 Census total population that is between 5 and 17 years old (square root transformed)	2016 PDB Core Variable

pct_nh_blk_alone_cen__logit	The percentage of the 2010 Census total population that indicate no Hispanic origin and their only race as "Black, African Am., or Negro" or report entries such as African American, Kenyan, Nigerian, or Haitian (logit transformed)	2016 PDB Core Variable
pct_sngl_prns_hhd_cen__sqrt	The percentage of all 2010 Census occupied housing units where a householder lives alone (square root transformed)	2016 PDB Core Variable
pct_not_hs_grad_acs__sqrt	The percentage of the ACS population aged 25 years and over that are not high school graduates and have not received a diploma or the equivalent (square root transformed)	2016 PDB Core Variable
med_house_value_acs__log	Median of ACS respondents' house value estimates for the tract (log transformed)	2016 PDB Core Variable
pct_othr_lang_acs__sqrt	The percentage of the ACS population aged 5 years and over that speaks a language other than English at home (square root transformed)	2016 PDB Core Variable
pct_crowd_occp_u_acs__sqrt	The percentage of ACS occupied housing units that have more than 1.01 persons per room (square root transformed)	2016 PDB Core Variable
pct_no_ph_srvc_acs__sqrt	The percentage of ACS occupied housing units that do not have a working telephone and available service (square root transformed)	2016 PDB Core Variable
pct_mobile_homes_acs	The percentage of all ACS housing units that are considered mobile homes	2016 PDB Core Variable
urban	The percentage of the tract’s land area occupied by urban blocks	2016 PDB Core Variable
timing_flag	The number of days elapsed for tract to reach 50% response rate	2010 Census Timing Flags
access	The noise-injected percentage of housing units with internet access	ACS 4 Year Internet Access Measure
nonmobile	The noise-injected percentage of housing units with non-mobile internet access	ACS 4 Year Internet

		Access Measure
2020_psrs	The tract's noise-injected 2020 Predicted Self-Response Score	2020 LRS/IPSR
2020_ipsr	The tract's noise-injected 2020 Internet Proportion of Self-Response Score	2020 LRS/IPSR

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/,
2016 Census Planning Database. https://www.census.gov/research/data/planning_database/2016/,
2010 Census Take 10 Data.

Table 19. Factor Loadings of Segmentation Variables

2020 Segmentation Variable Name	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
pct_renter_occp_hu_cen__logit	-0.25	-0.45	0.34	0.54	0.22	0.23	0.13
pct_pop_18_24_cen__sqrt	-0.10	-0.07	0.09	0.87	0.07	-0.11	0.07
pct_female_no_hb_cen__sqrt	-0.60	0.22	0.17	0.11	0.61	0.23	0.17
pct_nh_white_alone_cen__logit	0.28	-0.13	-0.52	-0.19	-0.52	-0.16	-0.40
pct_pop_65plus_cen__sqrt	-0.08	-0.36	-0.18	-0.59	-0.01	-0.55	-0.18
pct_rel_under_6_cen__sqrt	-0.28	0.25	0.33	0.44	0.21	0.56	0.00
pct_males_cen	0.02	0.03	0.10	0.15	-0.46	0.17	0.07
pct_mrdcple_hhd_cen	0.42	0.62	-0.06	-0.35	-0.45	-0.18	-0.13
pct_pop_25_44_cen__sqrt	0.13	-0.14	0.25	0.16	0.04	0.83	0.04
pct_vacant_units_cen__log	-0.55	-0.34	-0.12	-0.02	-0.24	-0.02	0.31
pct_college_acs__logit	0.84	-0.21	-0.15	-0.03	0.04	-0.06	0.13
med_hhd_inc_acs__log	0.85	0.25	-0.06	-0.32	-0.11	-0.01	0.09
pct_pop_45_64_cen__sqrt	0.13	0.03	-0.27	-0.75	-0.13	-0.14	0.03
avg_tot_prns_in_hhd_cen__log	-0.05	0.88	0.43	0.05	0.01	0.06	0.05
pct_hhd_moved_in_acs__sqrt	-0.06	-0.36	0.14	0.56	0.09	0.27	0.05
pct_hispanic_cen__logit	-0.03	0.12	0.81	0.18	0.09	0.14	0.01
pct_single_unit_acs__logit	0.25	0.57	-0.28	-0.25	-0.04	-0.16	-0.08
popdensity_cen__log	0.26	-0.15	0.43	0.24	0.72	0.12	-0.05
pct_prs_blw_pov_lev_acs__sqrt	-0.72	-0.16	0.19	0.41	0.15	0.05	0.06
pct_diff_hu_1yr_ago_acs__sqrt	-0.08	-0.37	0.03	0.66	0.04	0.20	0.03
pct_pop_5_17_cen__sqrt	-0.17	0.76	0.07	-0.22	0.10	0.24	-0.04
pct_nh_blk_alone_cen__logit	-0.27	-0.02	-0.07	0.20	0.60	0.22	0.37
pct_sngl_prns_hhd_cen__sqrt	-0.18	-0.90	-0.17	0.06	0.19	0.06	-0.03
pct_not_hs_grad_acs__sqrt	-0.75	0.12	0.48	0.07	0.04	0.09	-0.04
med_house_value_acs__log	0.75	-0.03	0.29	-0.11	0.05	-0.09	0.20
pct_othr_lang_acs__sqrt	0.11	0.05	0.91	0.13	0.12	0.09	0.04
pct_crowd_occp_u_acs__sqrt	-0.31	0.13	0.67	0.16	0.02	0.15	0.06
pct_no_ph_srvc_acs__sqrt	-0.40	-0.19	0.06	0.18	0.01	0.15	0.00
pct_mobile_homes_acs	-0.39	0.05	-0.09	-0.14	-0.44	-0.05	-0.03
urban	0.26	-0.14	0.31	0.21	0.70	0.09	-0.05

timing_flag	-0.26	-0.08	0.31	0.21	0.15	0.13	0.35
access	0.84	0.16	-0.01	0.04	0.08	0.12	-0.25
nonmobile	0.87	0.13	-0.02	-0.02	0.10	0.09	-0.23
2020_psrs	0.72	0.06	-0.39	-0.19	0.00	-0.08	-0.53
2020_ipsr	0.69	-0.02	-0.01	0.05	-0.03	0.10	-0.15

Source: American Community Survey. https://www.census.gov/programs-surveys/acs/,
2016 Census Planning Database. https://www.census.gov/research/data/planning_database/2016/,
2010 Census Take 10 Data.

B.6 Clustering Algorithms

Clustering algorithms produced both the mindsets and tract segments described in Sections 5 and 6 of this report. The CRAT team used the k-means and Ward's method clustering algorithms. The sections below describe these algorithms, as well as details about how the CRAT team used them to produce mindsets and tract segments.

B.6.1.1 K-means

The k-means algorithm is a widely used partitional clustering algorithm that assigns all observations to one of *k* clusters (Reddy and Vinzamuri, 2014). The number of clusters, *k*, must be defined before the algorithm is run. The CRAT team examined solutions to k-means algorithms for values of *k* that ranged from 6 to 12.

Because k-means begins by randomly selecting centroids, it can converge to one of a set of local minimums rather than the global minimum. If a particular solution reflects a local rather than a global minimum, then re-running the algorithm with the same number of clusters on the same data can produce a different result (Steinley, 2006). To avoid using an unstable solution, the CRAT team ran the k-means algorithm for a particular value of *k* a total of 20 times and selected the solution with the lowest mean square error.

B.6.1.2 Ward's Method

The agglomerative hierarchical clustering algorithm builds a sequence of nested clustering solutions, with the number of clusters ranging from one cluster per observation to a single cluster of all observations. Ward's is one of multiple linkage methods available for the identification of which clusters to combine. Some other linkage methods include single-link clustering, which combines the two clusters whose nearest members are closest, and complete-link clustering, which measures the similarity between clusters as the farthest distance between points in each cluster.

Ward's linkage method links clusters together based on the magnitude by which the sum of squared errors of the linked clusters exceeds the total sum of squared errors of each separate cluster. This linkage method aims to minimize the increase in intracluster distance at each step and it tends to produce compact, equal-size clusters. The CRAT team employed hierarchical clustering with Ward's linkage method to find solutions ranging from six to 12 clusters.