Diving into the U.S. Census Bureau Planning Database and ROAM Application: Tools for Survey and Census Planning

Kathleen Kephart

Center for Behavioral Science Methods (ADRM) Fed CASIC April 13th, 2021



Overview

- The Planning Database: Background, Introduction, and Examples
- What variables may be of use to me?
- Live Demo of the Response Outreach Area Mapper (ROAM)
- What can you expect from the 2021 Planning Database?



Where Can I Find ACS Estimates?

• Explore Census Data Tool - data.census.gov (after July 1, 2019)

Planning Database:

census.gov/topics/research/guidance/planning-databases.html



Levels of Geography Available

- The ACS sample is aggregated over 5 years to produce estimates at census tract and block group levels
- A census tract is a subset of a county
 - More than 74,000 in the U.S.
 - Generally have a population size between 1,200 and 8,000; optimum size is about 4,000 people
- A block group is a subset of a census tract
 - More than 220,000 in the U.S.
 - Generally have a population size between 600 and 3,000



Why do we have a Planning Database (PDB)?

- To plan for field resources and to create tailored communication and partnership campaigns
- 1990: Hard to Count Score (HTC) and an early internal PDB
 - Households in each census tract were assigned a score
 - The higher the score, the harder to count
 - Field Division used the score and internal PDB to make hiring decisions and resource allocations
- 2000 and 2010: Continued hiring and resource decisions made with it.
 Additional staff across the country used the score to identify areas that needed extra outreach, engagement, and education
- 2020 Census: HTC upgraded to Low Response Score, and it is now free and publicly available to internal and external stakeholders



What is in the Planning Database?

- "Greatest hits" of ACS 5-year estimates and 2010 Census variables
- Types of variables
 - Population: gender, age, education, poverty
 - Household: language, relationship, income
 - Housing unit: tenure, number of units
 - Census operational: mailout/mailback, bilingual
- We calculate the percentage variables for each estimate
- Available at census tract and block group level
- Available as a comma-delimited file (CSV) or through the Census API



Broad Scope of Uses

- Useful for:
 - Identifying areas with likely low survey response rates
 - Stratifying small areas
 - Creating thematic maps
 - Enhancing reports with population metrics

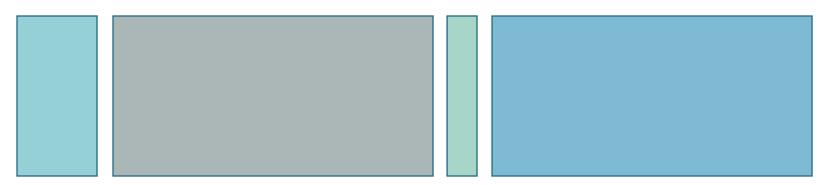


Why Use the Planning Database?

- Easier to download than full ACS Summary Files
- Contains most popular ACS 5-year census tract & block group aggregated estimates
- Estimates are matched to 2010 Census estimates for each geography
- Contains this decade's hard to count score—the Low Response Score (LRS)
- Updated annually with the latest ACS data
- The ROAM application makes the PDB even more accessible



The Structure



Geography Identifiers

- GIDBG (12 chars) = State (2 chars) + County (3 chars) + Tract (6 chars) + Block Group (1 char)
- GIDTR (11 chars) = State (2 chars) + County (3 chars) + Tract (6 chars)

Demographic, Socioeconomic, and Housing data

- Order of variables is consistent. Census data first, followed by ACS estimates and ACS MOEs
- For example, Males_CEN_2010, Males_ACS_13_17, Males_ACSMOE_13_17

Census Operational data including Mail Return Rate and Low Response Score

Percentages and MOE Percentages. Listed in the same order as their respective estimate

- Variables identified with 'pct_' added to their variable name
- For example, pct_Males_CEN_2010, pct_Males_ACS_13_17, pct_Males_ACSMOE_13_17



Ways to Use The Planning Database



Area Demographics: District of Columbia

• 672,391 people live in 179 tracts in the District of Columbia

	DC	United States
Male to female ratio	0.90	0.97
Population under 5 years old	6.5%	6.2%
Population that identifies as Hispanic	10.7%	17.6%
Population that was not born in the US	14.0%	13.4%

Source: 2013-2017 ACS 5-year Estimates



2016 Census Test: Harris County, Texas Demographics

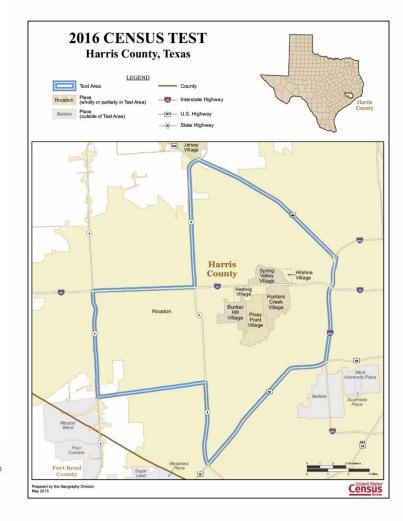
• 484,358 people live in 292 block groups in the test site

	Harris County, TX	United States
Households where no one over 14 speaks English "very well"	14.8%	4.6%
Population 18-24 years old	9.4%	10.0%
Renter Occupied Units	60.9%	35.1%
Population 25 and over, with less than a HS diploma	19.1%	13.9%

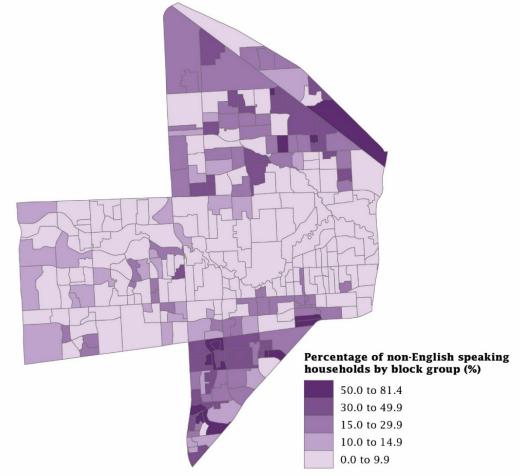
Source: 2009-2013 ACS 5-Year Estimates



2016 Census Test: Harris County, Texas



Households Where No One Age 14 and Over Speaks English Only or Speaks English "Very Well"





Source: U.S. Census Bureau 2015 Planning Database, 2009-2013 5-year American Community Survey (ACS) estimates

Non-English Speaking Households

- What if you want to identify areas that may need support for a language other than English?
- Find block groups in the area that have a high percentage of housing units where no one over the age of 14 speaks English "very well"
- What language is spoken in these block groups?



2016 Census Test:

Non-English Speaking Block Groups in Harris County, TX

Rank	BG	No one speaks English "very well"	Spanish	Asian/Pacific Islander	Other
1	4327012	81.4% (14.3)	81.4% (14.0)	0% (2.1)	0% (2.1)
2	4330012	77.2% (13.4)	73.4% (13.5)	3.8% (4.1)	0% (2.3)
3	4327011	72.5% (11.1)	72.5% (10.9)	0% (1.6)	0% (1.6)
4	4335012	69.3% (10.9)	66.1% (10.7)	0% (1.7)	3.2% (4.8)
5	5214001	69.3% (21.1)	69.3% (20.6)	0% (3.7)	0% (3.7)



Source: 2009-2013 ACS 5-year Estimates

What is the Low Response Score (LRS)?

- LRS = predicted level of Census self <u>non-response</u> at the census tract or block group level
- Based on Ordinary Least Squares (OLS) regression of two dozen metrics from the PDB upon the Census 2010 mail return rate
- Values range from 0-100 percent
- For example, if LRS=25, we are estimating that 25 percent of households in that census tract will not self-respond to the 2020 Census
- LRS is updated yearly using the latest ACS 5-year estimates



LRS OLS Linear Model (Block Group edition)

	Coef	Sig		Coef	Sig
(Intercept)	10.29	***	Renter occupied units	1.08	***
Ages 18-24	0.64	***	Female head, no husband	0.58	***
Non-Hispanic White	-0.77	***	Ages 65+	-1.21	***
Related child <6	0.46	***	Males	0.09	***
Married family households	-0.12	***	Ages 25-44	-0.06	
Vacant units	1.08	***	College graduates	-0.32	***
Median household income	0.24	***	Ages 45-64	-0.08	*
Persons per household	3.44	***	Moved in 2005-2009	0.09	***
Hispanic Any Race	0.41	***	Single unit structures	-0.52	***
Population Density	-0.40	***	Below poverty	0.11	***
Different HU 1 year ago	-0.12	***	Ages 5-17	0.17	***
Non- Hispanic Black	-0.04	**	Single person households	-0.24	***
Not high school grad	-0.06	***	Median house value	0.71	***

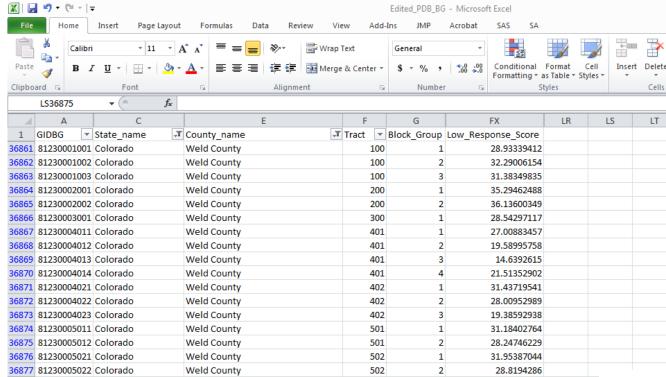
Sig: *** p < .001; ** $.001 \le p < .01$; * $.01 \le p < .05$ R-

squared: 56.10%, n = 217,417

Main Effects only, no interaction terms



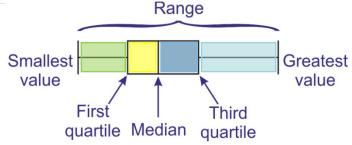
Identifying LRS Groups



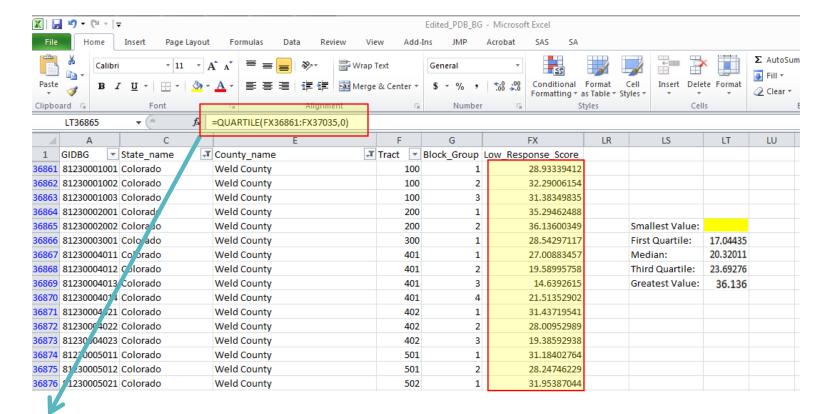
Here I have hidden all columns except for some geo locating variables and the Low Response Score.

I have also applied a filter to only show data from Weld County, CO.

Use Excel to help us locate and identify those block groups with a Low Response Scores that fall within the quartile groups.







=QUARTILE(Starting Cell:Ending Cell,0) → Smallest Value

=QUARTILE(Starting Cell:Ending Cell,1) → First Quartile

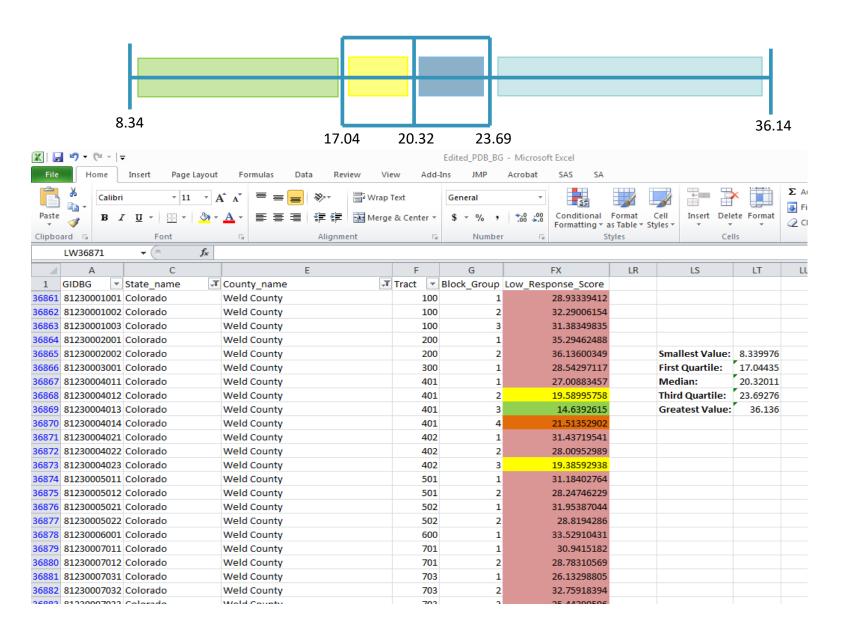
=QUARTILE(Starting_Cell:Ending_Cell,2) → Median

=QUARTILE(Starting_Cell:Ending_Cell,3) → Third Quartile

=QUARTILE(Starting_Cell:Ending_Cell,4) → Greatest Value

Tip: When selecting your starting and ending cells, after selecting your starting cell hold shift and scroll to your ending cell and select. This will highlight all cells in between.







Limitations of the Low Response Score

- Only considers <u>mail</u> self-response mode 2020 Census will offer internet, phone, AND mail
- Some census tracts have small number of housing units in mailback universe (e.g., Indian reservations, very rural areas)
- LRS cannot be generated for Puerto Rico or select census tracts with post-2010 boundary changes
- If LRS is extremely high, take a closer look
 - There are not currently margins of error (MOEs) for LRS



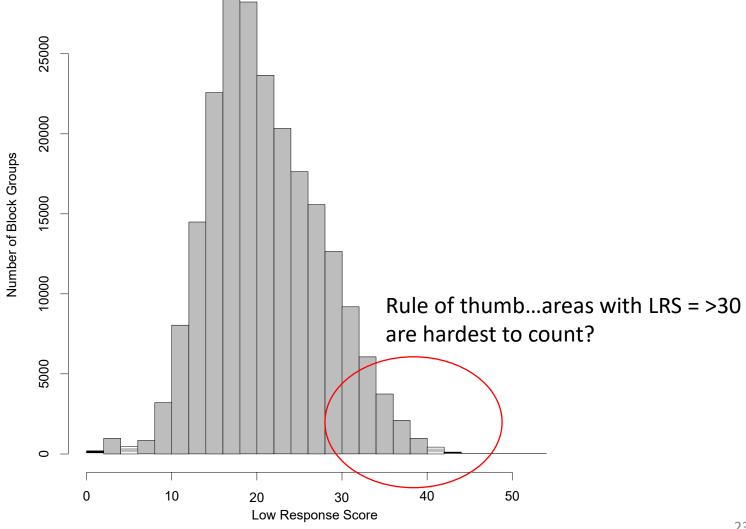
Full Discussion of LRS Methodology:

Erdman, C. and N. Bates (2017). The Low Response Score (LRS): "A Metric to Locate, Predict, and Manage Hard-to-Survey Populations", Public Opinion Quarterly, Volume 81, Issue 1, 1 March 2017, pp. 144-156.

https://www.census.gov/research/data/planning database/2018/docs/erdman bates 2017.pdf



National Distribution of the LRS





Local Application of the LRS

- What constitutes a "high" LRS depends
- Are you concerned with a particular State, Census Region, Place, or County?
- If yes, extract census tracts within that geographic area and produce custom LRS distribution
- What score is one or two standard deviations above the mean?
- Identify census tracts matching your cutoff



ACS Self-Response Rate

- 5-year self response for internet and mail to the ACS, available only at the tract level
- This is the percentage of households in the tract that responded by internet or mail to the ACS, aggregated over 5 years.
- It is ONLY available on the PDB, not on data.census.gov
- Important things to know
 - These rates do not include group quarters
 - There are no ACS self-response rates for Puerto Rico at this time
 - If tracts do not have a minimum number of households they will not have an ACS self-response rate

LRS Correlation with ACS and 2018 Census Test self-response (tract level)

	R ² for propensity models			
	2010	ACS 2013-	2018 End-to-	
Independent variables	Census	2016*	End Census	
	mail return	self-response	Test* return	
	rate	rate	rate	
	n=71,657	n=71,454	N=141	
Low Response Score				
(LRS)	0.59	0.64	0.87	



^{*}Internet and mail responses combined

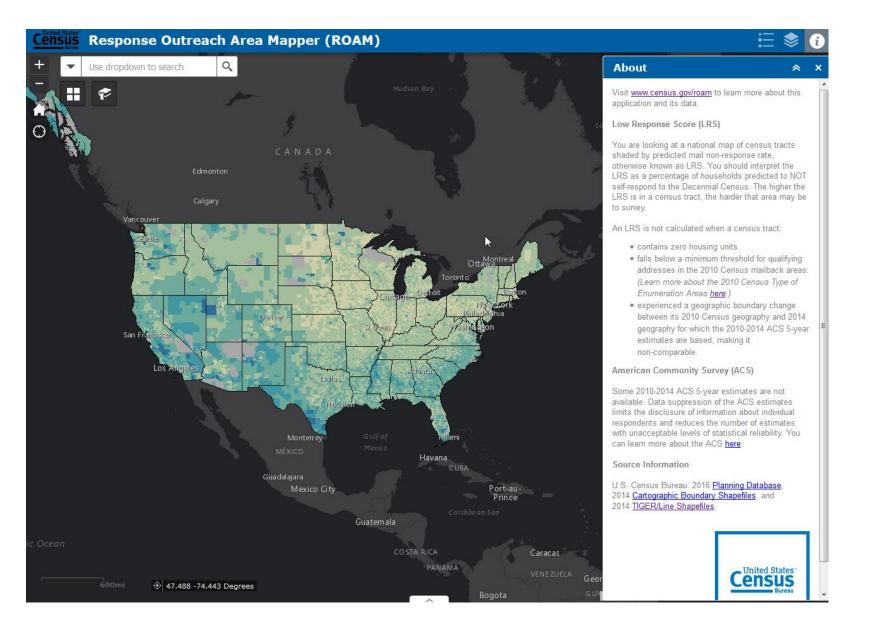
What is the Response Outreach Area Mapper (ROAM)?

- A web mapping application that provides a map and data table interface for users to identify and learn about hard-to-count census tracts
- Accessible through a web browser (no download required)
- Hosted by the Census Bureau



ROAM Demo

www.census.gov/roam





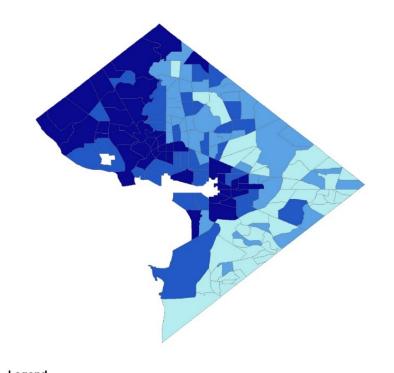
New in 2019 PDB and ROAM

- ACS 5-year Internet and Technology variables are now available at the census tract level
 - HHDs with broadband access
 - HHDs with smartphone only access, etc.
- We are adding several variables about children to the PDB at the census tract level
 - 3-4 year old school enrollment
 - HHDs with child under 5 present, etc.
- ACS median age available in the PDB at census tract and block-group levels
- 2020 Census Communications Campaign audience segmentation variable (census tract-level)



2013-2017 ACS 5-Year Census Tract Self-Response Rate

ACS 5-year 2013-2017 Tract Self-Response Rate in D.C.







Tract Segmentation Goals and Applications

The goals of the segmentation are to:

- Provide an overarching framework for understanding the country
- Use geography to bring together behavioral, demographic, attitudinal, and media usage data for campaign planning
- Simplify complex data by identifying key shared characteristics

Segmentation will inform:

- Overall program strategy
- Planning for advertising, communications, and partnership activities
- Creative development
- Campaign optimization and tracking during enumeration





Responsive Suburbia

71%

Of the U.S. Population

24%

Predicted Self-Response

- High predicted rate of response, with a high percentage of that response coming online.
- Found in suburban neighborhoods of singlefamily homes.
- High % college educated, high % married, and high median household incomes.



Main Street Middle

67%

Predicted Self-Response



- High predicted rate of response, with an above-average percentage of that response coming online.
- Found in small towns and less densely populated areas surrounding urban centers.
- · Low diversity and a higher % 65 or older than the national average.



Country Roads

60%

Predicted Self-Response



- Slightly below-average predicted rate of response, with a below-average percentage of that response coming online.
- Found in rural areas predominantly in the eastern United States, surrounding small towns and outside the suburbs of major cities.
- High % owner-occupied housing, low % college educated, and below-average median household incomes.



Downtown Dynamic

59%

Predicted Self-Response



- Slightly below-average predicted rate of response, with a high percentage of that response coming online.
- Found in densely populated metro centers.
- High % college educated, above-average % foreign-born, high % 25-44 compared to the nation as a whole, and high median household incomes.

Tract Segmentation Goals and **Applications**



Student and Military Communities

56% Predicted Self-Response

2%

Of the U.S. Population

- Below-average predicted rate of response, with a high percentage of that response coming online.
- · Found in communities around college campuses or military bases.
- A majority 18-24, high % college educated, and high % renter-occupied housing.



Sparse Spaces

49%

Predicted Self-Response



- Below-average predicted rate of response, with below-average internet response.
- Found in rural areas predominantly in the western United States, Appalachia, northern Maine, and Michigan's Upper Peninsula.
- · High % owner-occupied housing and belowaverage levels of internet access.

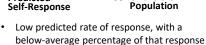


Multicultural Mosaic

45%

Predicted Self-Response

coming online.



Of the U.S.

14%

- Found in California's Central Valley and parts of New Mexico, Texas, Florida, as well as concentrations in urban areas.
- High % foreign-born, low % college educated, and majority Hispanic.



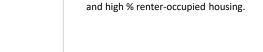
Rural Delta and Urban Enclaves

43%

Predicted Self-Response



- Low predicted rate of response, with the lowest percentage of that response coming
- Found in rural parts of the southeastern United States, as well as concentrations in urban areas.
- Low % college educated, low median household incomes, below-average levels of internet access, and majority non-Hispanic African American.









Multicultural Mosaic: Demographics

Where are they?

Tracts in this segment can be found in California's Central Valley and parts of New Mexico, Texas, and Florida, with additional concentrations in urban areas.

Who are they?

The tracts in this segment have relatively high concentrations of foreign-born residents, as well as low percentages of college-educated residents. A majority of the people in this segment are Hispanic.

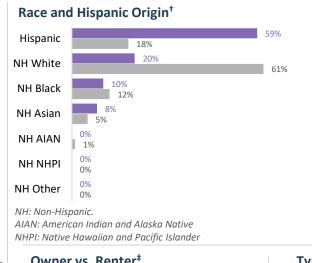
How will they respond?

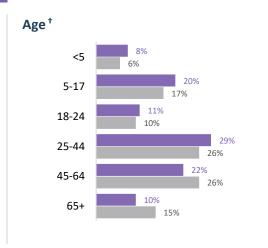
On average, this segment is predicted to respond at a low rate, with a below-average percentage of that response coming online.

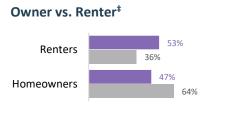
14% of U.S. Population

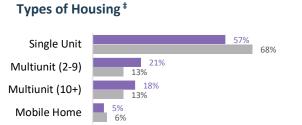


National: 77%









^{† -} Population Average; ‡ - Household Average
Source: 2016 5-year ACS estimates unless otherwise marked by * (2020 ICC Modeled Scores, DRB# CBDRB-FY18-311)
or & (ACS data from 2013-2017, DRB# CBDRB-FY18-311).

National: 66%



Multicultural Mosaic: Demographics and Mindsets

Where are they?

Tracts in this segment can be found in California's Central Valley and parts of New Mexico, Texas, and Florida, with additional concentrations in urban areas.

Who are they?

The tracts in this segment have relatively high concentrations of foreign-born residents, as well as low percentages of college-educated residents. A majority of the people in this segment are Hispanic.

How will they respond?

On average, this segment is predicted to respond at a low rate, with a below-average percentage of that response coming online.

14% of U.S. Population

Multicultural Mosaic



Foreign-Born[†]



34%

National: 13%

Non-English-Speaking [‡]



18%

National: 5%

College-Educated †



15%

National: 30%

Median HH Income ‡



\$45,256National: \$60,835

Married Couple HH[‡]



45%

National: 48%

HH With Children[‡]



43%

National: 32%

Family HH With Related Children Under 6 [‡]



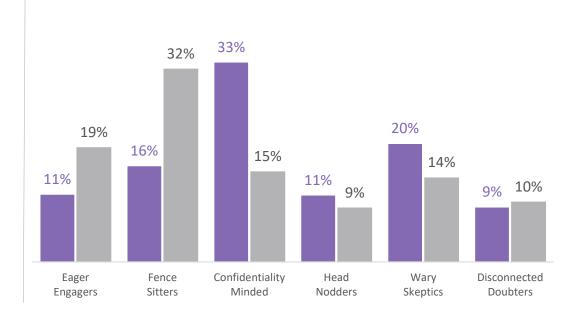
28%National: 21%

Moved in the Last Year ‡



13% National: 37%

Mindset Composition[†]



^{† -} Population Average; ‡ - Household Average



Multicultural Mosaic: Media Usage

Where are they?

Tracts in this segment can be found in California's Central Valley and parts of New Mexico, Texas, and Florida, with additional concentrations in urban areas.

Who are they?

The tracts in this segment have relatively high concentrations of foreign-born residents, as well as low percentages of college-educated residents. A majority of the people in this segment are Hispanic.

How will they respond?

On average, this segment is predicted to respond at a low rate, with a below-average percentage of that response coming online.

14% of U.S. Population

Compared to the national average, the average person in this segment...



Reads a **similar** number of **newspapers** in a 28-day period.



Listens to a **similar** number of half-hours of **radio** in a week.



Reads a **similar** number of **magazine** issues in a month.



Views **similar** number of half-hours of **TV** in a week.



Drives 13% **fewer** miles in a car or truck in a 7-day period.



Uses a **similar** number of hours of **internet** in a week.

Questions?

census.PDB.questions@census.gov

