

# Hard-to-Survey (HTS) Populations and the US Decennial Census

*26<sup>th</sup> Annual Morris Hansen Lecture*

Washington DC

Nancy Bates

Senior Researcher for Survey Methodology

U.S. Census Bureau

- Who are hard to survey?
- Social Marketing campaigns in the US Census
  - Hard to Survey Segments
  - Evidence of campaign on HTS segments
- New HTS metric
- Digital media and HTS populations

# *Hard-to-Survey: Frameworks*

Hard to survey according to *survey process*  
(Tourangeau, 2014)

- Hard to sample,
- locate,
- contact,
- persuade, and/or
- interview

# *HTS groups*

Undercounted in 2010 Census (Mulry, 2014)

- Ethnic minorities (Blacks, Hispanics, AIAN)
- Males age 18-29
- Non-nuclear family members

High level of effort/proxy (Walker et al., 2012)

- Hispanic-headed households
- Black-headed households
- Renter households

# *HTS groups*

## Undercounted young children (0-4)

- Mothers (Konicki and Griffin, 2016; Dolson 2013)
  - Young (15-24)
  - Unmarried
  - Not living in own household
- Black and Hispanic; limited English proficiency; < HS education; unemployment (O'Hare, 2014)

# *Methods to count HTS*

- Census special operations/forms
  - Shelter & Service-based enumeration
  - Update/enumerate
  - Group quarters
  - Campsites/RV parks
  - “Be counted” forms

# *Methods to include HTS*

- Targeted mailouts:
  - Bi-lingual forms (Bentley, 2012)
  - Replacement forms (Hill, 2012)
  - Internet push vs Choice (Nichols, Horwitz, and Tancreto, 2013; Chesnut, 2015)

# *Methods to count HTS*

Innovations for 2020 Census (Blumerman, Bishop and Dinwiddie, 2016)

- Multiple modes of self response
- Response without unique ID code
- Online forms in multiple languages



# *Methods to include HTS: social marketing campaigns*

“The adaptation and adoption of **commercial marketing activities**, institutions, and processes as a means to **induce behavioral change** in a **targeted audience** on a temporary or permanent basis **to achieve a social goal**”

- S. Dann, 2010

# *Methods to include HTS: social marketing campaigns*

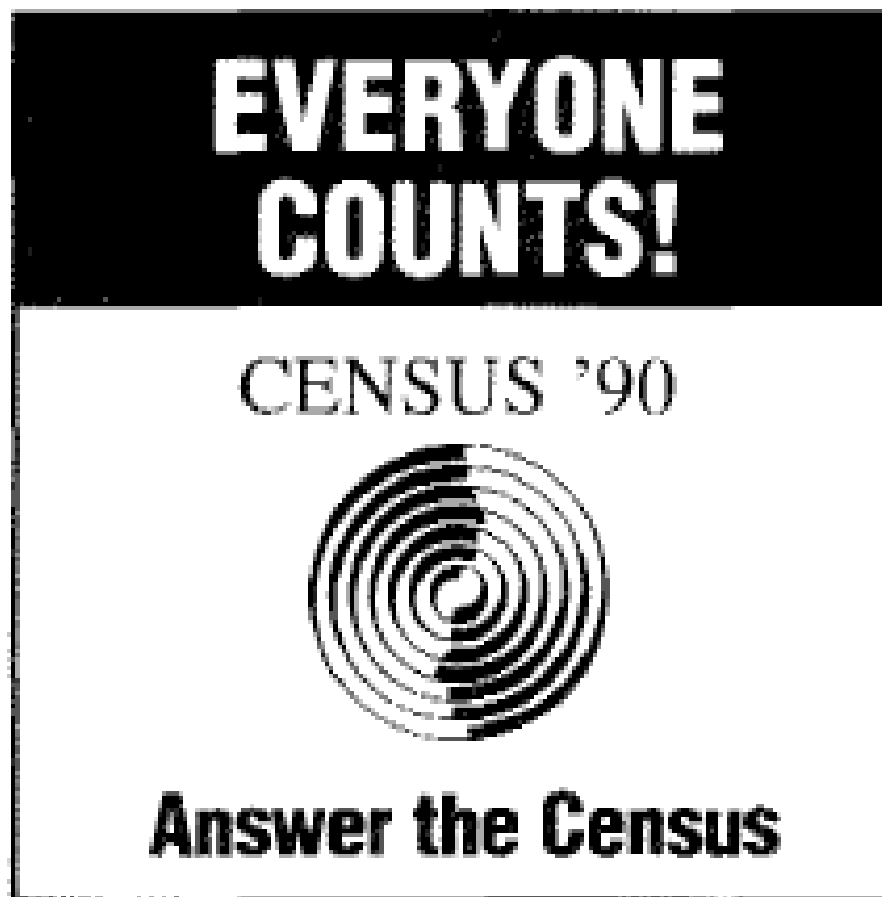
- Locate (trusted voices)
- Contact (direct mail, texts, SMS)
- Persuade (targeted messaging)
- Interview (click to complete)

# *Methods to include HTS: social marketing campaigns*

- Ad Council – Beginning with 1950 Census
- 1980 Census – PSAs
- 1990 Decennial Census - Partnered with Ogilvy and Mather
- Concentration on racial/ethnic minorities



# Census '90 Button





**LISTEN  
TO THE  
DRUM**

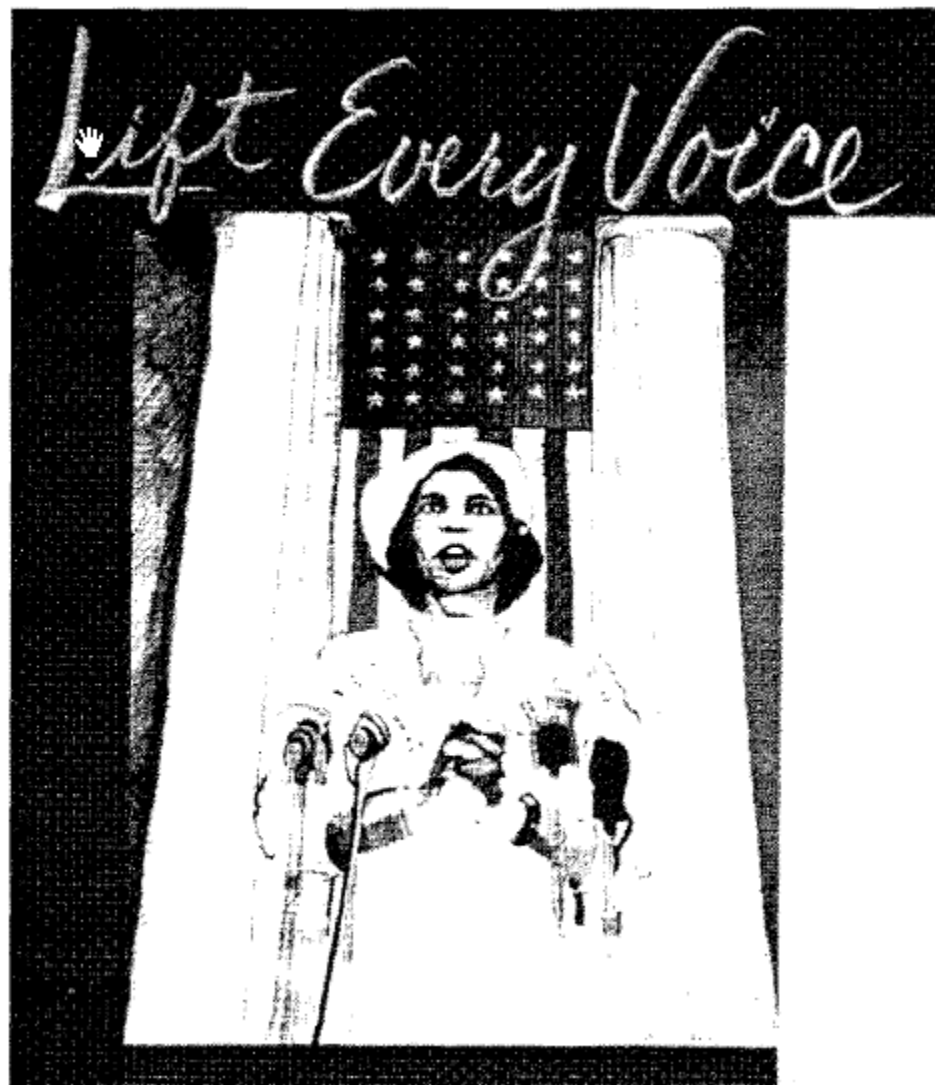
RAISE YOUR VOICE

IAIA

CENSUS '90



**Name Your Tribe.  
Answer The Census.**



# 1990 Census: a turning point?

- Budgeted for 70% self response
- By start of NRFU mail response was only 65%
- Spurred nonresponse research
- The agency decided to engage a paid campaign for 2000 Census

# 2000 and 2010 Campaigns

Census 2000 was first census to use *paid advertising*

- Engaged advertising contractors (2000 Young & Rubicam; 2010 DraftFCB)
  - Paid advertising (28 languages)
    - TV, radio, print, out-of-home, digital
    - 110 Million in 2000 and 167 Million in 2010
  - Earned media
  - National and local partnership program
    - Census in Schools
    - Faith-based outreach
    - Complete Count Committees



*Principle of social marketing?*

***Segment the target audience***

*(Adreason, 2002;French 2010)*

# *2010 Census segmentation*

- Census Planning Database
- Hard-to-Count (HTC) Score  
(Bruce, Robinson and Sanders, 2001)
- Census self-response data
- Augment with 2006 ACS data

# *Hard to Count Score Variables*

## **Housing Factors:**

- Percent vacant
- Percent Not Single Unit Structure
- Percent Renter Occupied
- Percent Crowded Occupancy

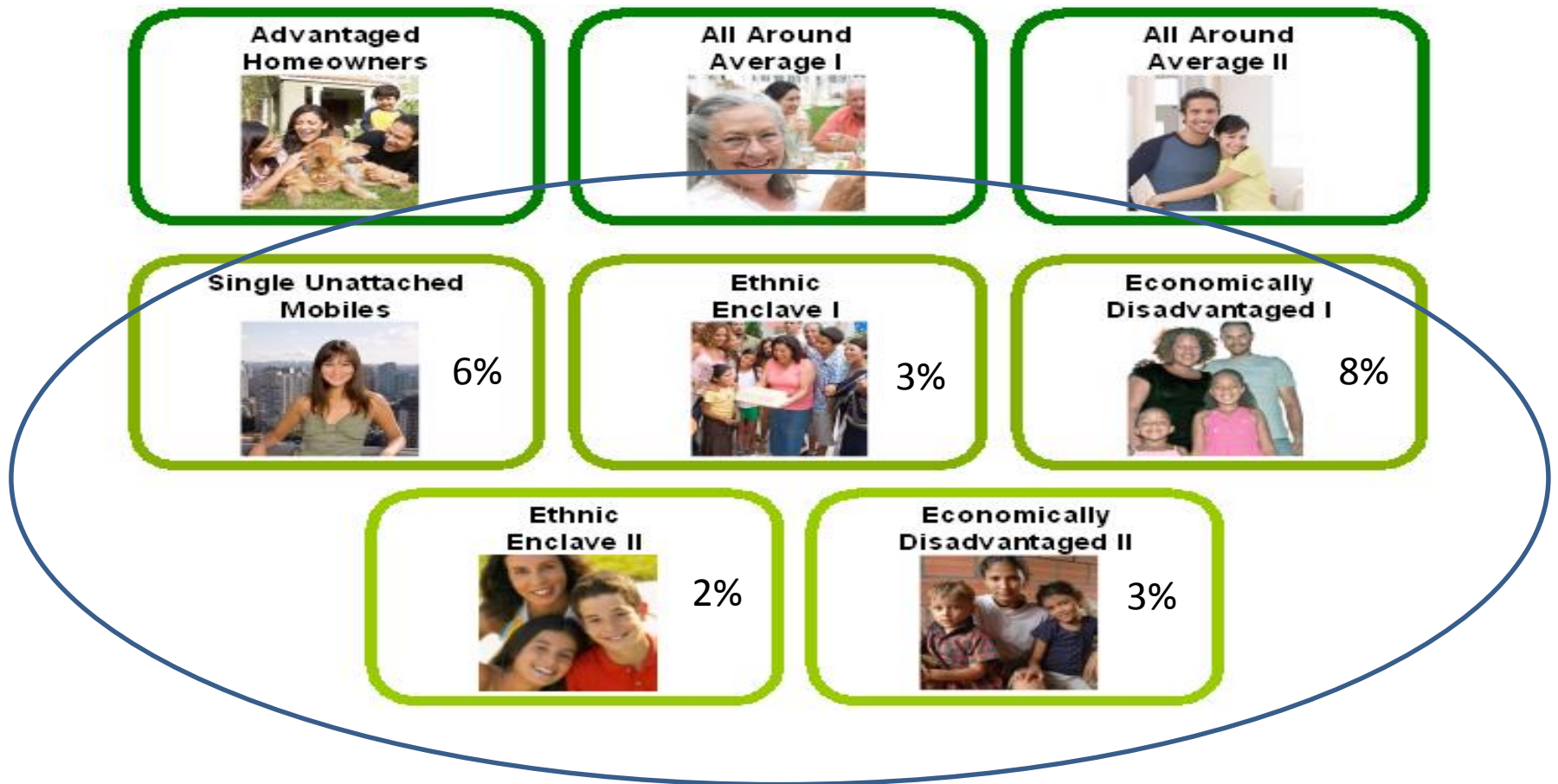
## **Social and Demographic:**

- Percent Not Husband/Wife Households
- Percent Household with no phone service
- Percent Not High School Graduate
- Percent with Public Assistance Income
- Percent Unemployed
- Percent below poverty level
- Percent Linguistically Isolate Households
- Percent Moved into Unit (1999-2000)

# *Segmentation Method*

- Mutually exclusive groupings of Census tracts
- k-means cluster analysis using Planning Database
- Input the 12 hard-to-count (HTC) score variables
- Yielded 8 clusters
  - Several iterations
  - Looking for constructs identified previously

# Geographic segmentation for 2010 Census social marketing campaign



Source: Bates and Mulry, 2012

# *Five segments considered HTS*

## Young/mobile/singles

renters; high educ; few children; urban; \$40K; racially diverse

## Economically Disadvantaged (owner skewed)

urban & rural; single mothers; \$26K; 49% Black; 1/3 live alone

## Economically Disadvantaged (renter skewed)

urban; female-headed; \$22K; 59% Black; 23% Hispanic

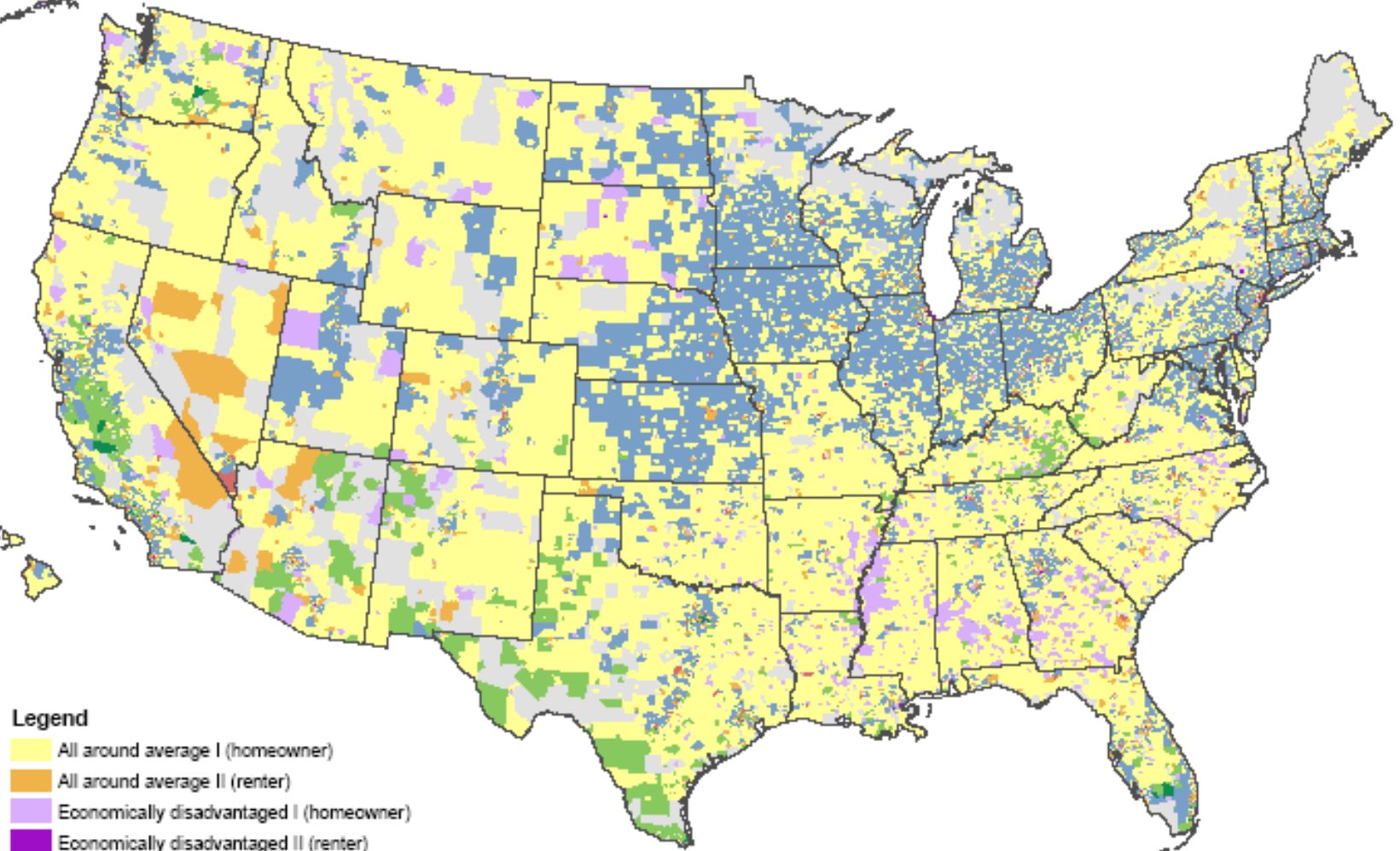
## Ethnic enclave (owner skewed)

43% foreign born; spousal; 50% w/children; \$35K; 61% Hispanic

## Ethnic enclave (renter skewed)

62% foreign born; low educ; younger; urban; \$32K; 59% Hispanic ;  
11% Asian; 34% limited English proficiency

# Segmentation Clusters for 2010 Integrated Communications Program



## Legend

- All around average I (homeowner)
- All around average II (renter)
- Economically disadvantaged I (homeowner)
- Economically disadvantaged II (renter)
- Ethnic enclave I (homeowner)
- Ethnic enclave II (renter)
- Single/mobile/unattached
- Advantaged homeowners
- Unclassified (Low population or > 35% age 65+)

*Apparently our HTS segments  
also apply to our neighbors to the  
North ...*



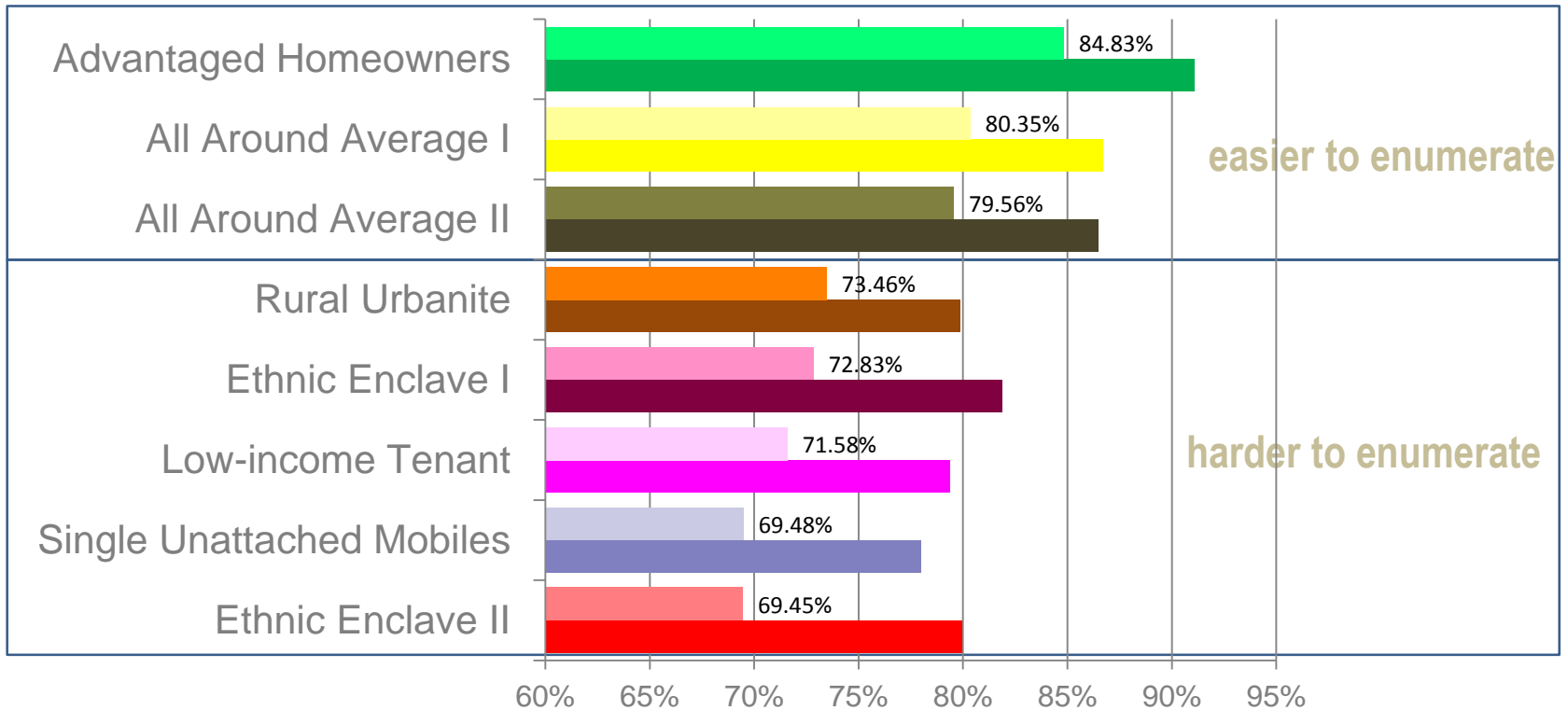


# CENSUS • RECENSEMENT

## Audience profiles

Self-response rates on June 2, 2011 (first line)

Final response rates on September 30, 2011 (second line)

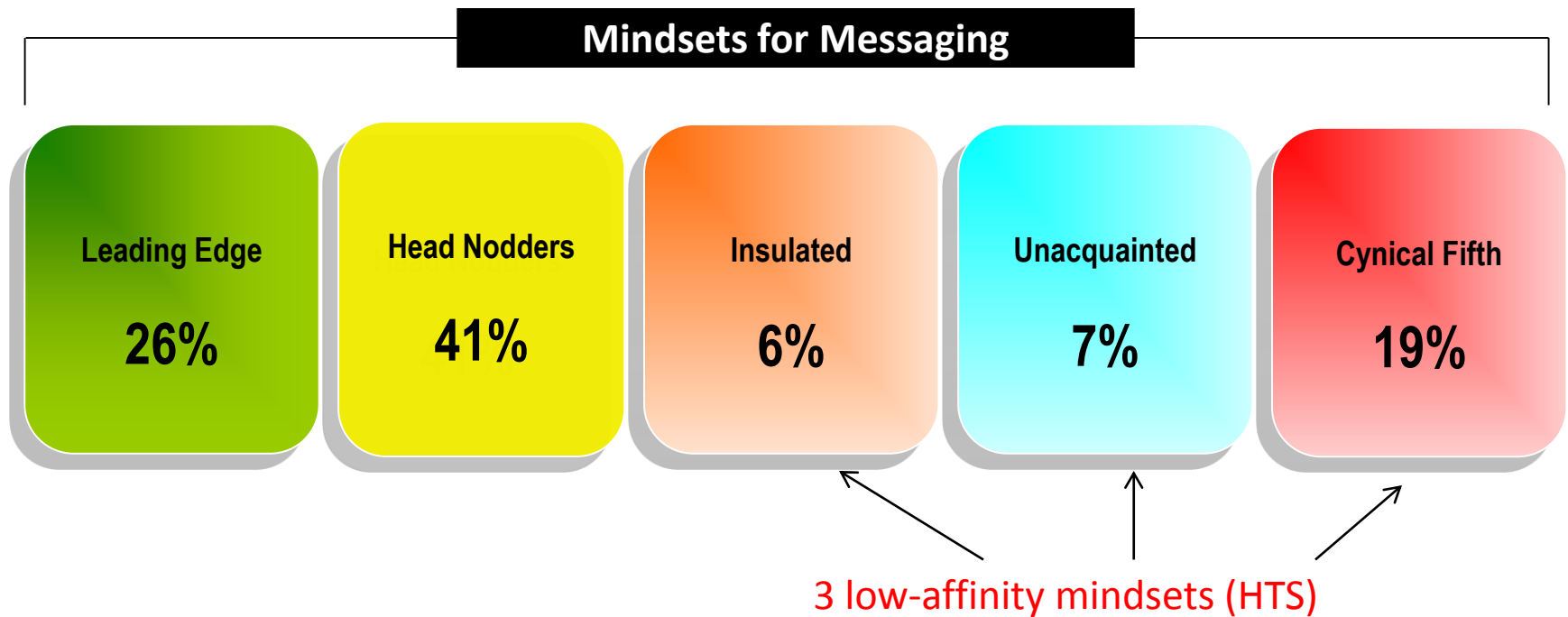


# *What about messaging to HTS?*

# *Census Barriers Attitudes and Motivator Survey (CBAMS)*

- Phone and in-person survey N=4000
- Oversampled HTS populations (e.g. limited English proficiency; high minority tracts)
- Measured:
  - Census familiarity and knowledge
  - Self-reported propensity to respond
  - Ranking of messages
  - Barriers and motivators to participation

# Attitudinal segmentation for 2010 Census campaign



Source: Census Bureau, CBAMS, 2008

# Insulated 6%

## Indifferent

Motivated by individual benefits of Census

88%

Female

68%

High school or less

56%



- Attached to their communications — higher on 10+ years in neighborhood
- Admittedly “don’t know” about Census — low familiarity
- Question impact of Census because they feel they haven’t seen results in their neighborhood
- Therefore, some softness of favorability, responsibility, trust and voice
- More interested in individual benefits of Census versus broader community benefits — probably due to their disadvantaged circumstance
- *Ethnic (Hispanic, Black, American Indian, other)*
- *Many don’t speak English at home (mostly Spanish)*
- *More female than any other segment*
- *Large pocket of older (65+) and widows — but all ages are reflected*
- *Less likely to have children*
- *Downscale: low income (\$25K or less, low education (more than half HS or less)*
- *Homeowners*

# Cynical Fifth 19% *Resistant*

Never see results of Census

63%

Info will be misused

46%

Favorable towards the Census

18%



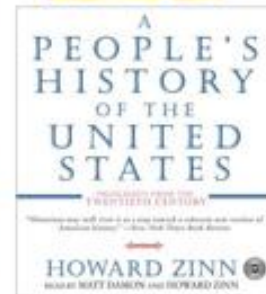
Leastern Conference



PRIVACY

Academy at New School University

11th Floor



- Highest claimed unfamiliarity with Census, but in reality, have relatively high knowledge
- Predisposition towards the Census is neutral to negative — alarmingly high belief they will never see results of Census in their neighborhood
- High belief information will be misused (or maybe not used) since no evidence to feel otherwise; also high belief that Census is an invasion of privacy
- Therefore low trust, high skepticism, but still some underlying sense of responsibility and belief that all should be counted
- Ideological
- Will have to convince them at another level, perhaps more emotional, gut
- Diverse (mirrors the population)
- Medium to high income
- Educated
- Male skew

# Unacquainted 7%

## Peripheral

- Never heard of the Census, even after aiding — know nothing about it at all
- Low community stakedness and civic participation — focusing on self for the moment — individual impact may have most leverage
- Least likely to indicate participation in the Census, but potential since no negative baggage
- Far less likely to vote — probably can't (non-citizens)
- Ethnic, majority minority (Hispanic, Asian, AI, Black)
- Almost half non-U.S. born
- Most speaking in language at home
- Male skew
- Younger or older
- Highest never married
- Largest household size (4+ people), most likely to have children in household
- Most downscale of all segments — least educated, lowest income
- Much more likely renters versus homeowners



Armed with audience segments and messaging mindsets, ad agency produced multitude of advertising pieces across various platforms...



WHAT CAN

1



QUESTIONS DO?

HELP BUILD BETTER COMMUNITIES



Better healthcare. Better schools. Better roads. All that is within your reach when you fill out the census form and mail it back when it comes in March. 10 simple questions. 10 minutes that will help decide how more than \$435 billion in federal funds get dispersed to your community.

For more information, go to [www.2010census.gov](http://www.2010census.gov).

▶▶ We move forward when you send it back.

United States  
Census  
2010  
IT'S IN OUR HANDS.

BOND

5

CENSUS 2010 - Miss Maybelle  
Online Conform :30 NFA  
Jan 5, 2010

[www.bondedit.com](http://www.bondedit.com)

# *Targeted ads*

- 35 % of African Americans recalled seeing Ms. Maybelle compared to 15% for Diverse Mass ad (Frank)
- 31% said targeted ad grabbed attention compared to 11% for DM ad
- 31% said gave reason to mail back compared to 11%

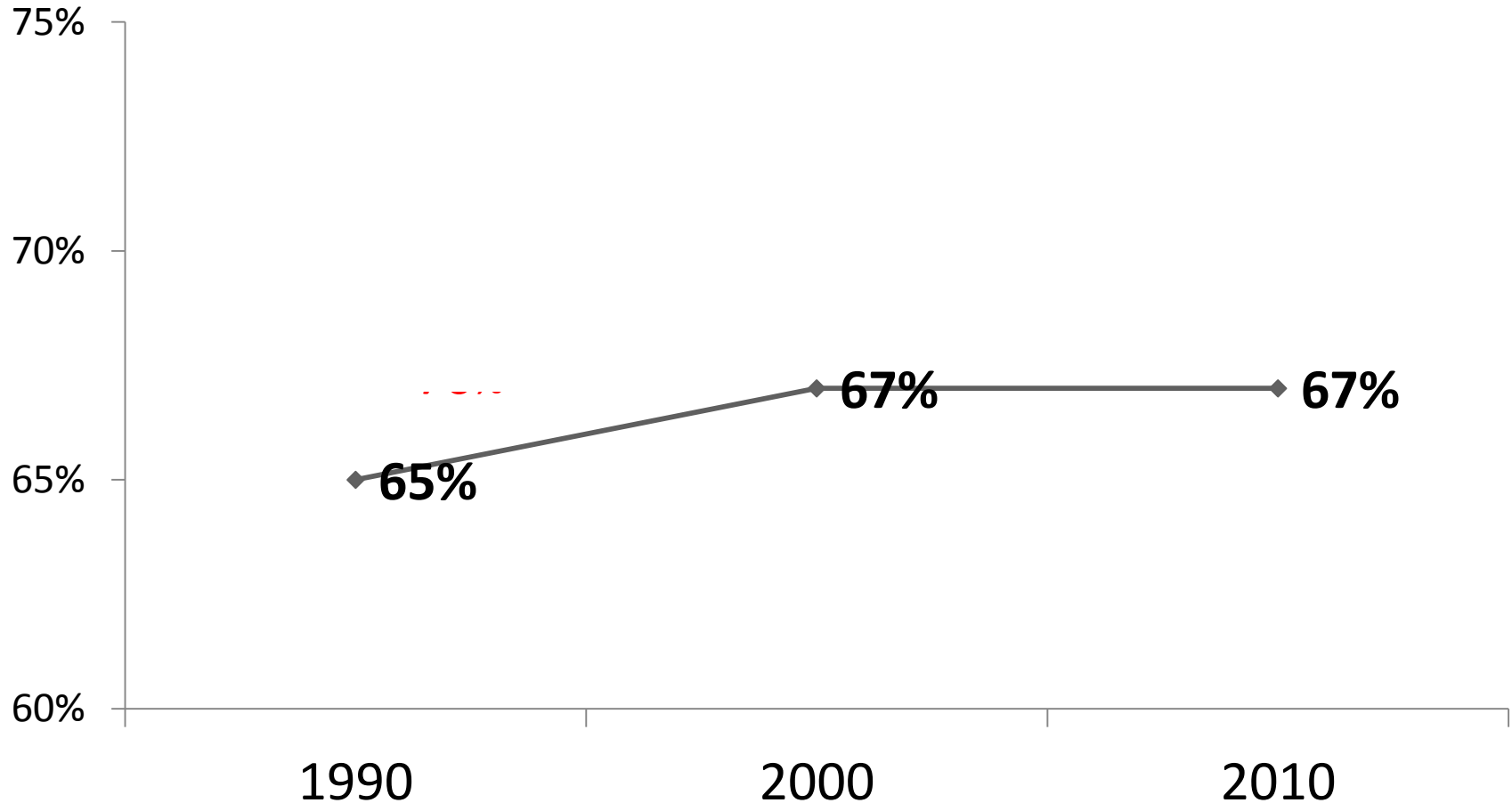
Conclusion: targeted ads resonated among target audiences

# *Breadth of 2010 Campaign*

- Between January and July 2010, **Census ranked among top 5 advertisers** behind McDonalds, Walmart and GEICO
- Average number of **ad exposures = 42**
- In fact, **ran out of minority media inventory** to purchase in some markets!

# *Do paid campaigns work?*

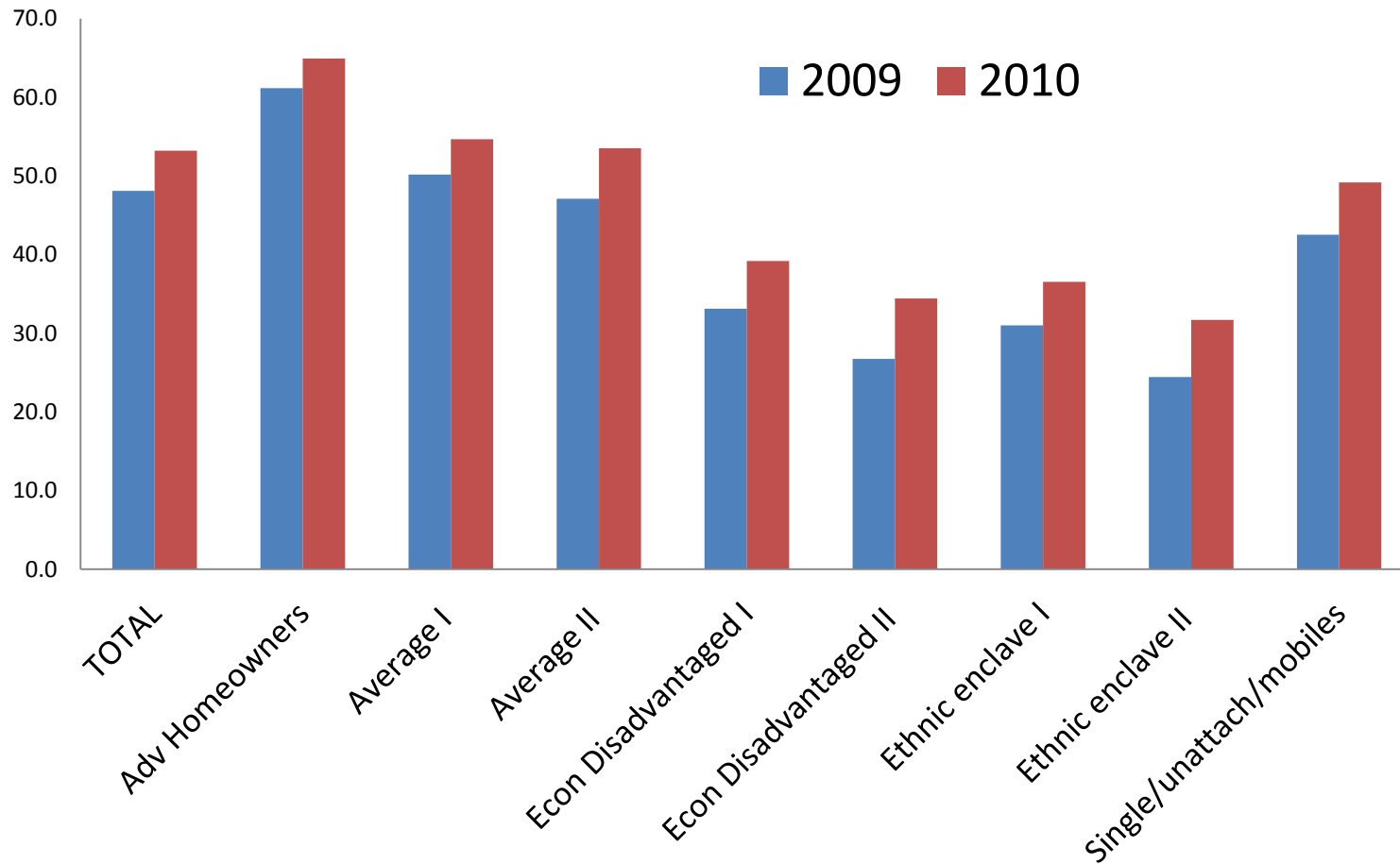
# Mail self-response rates



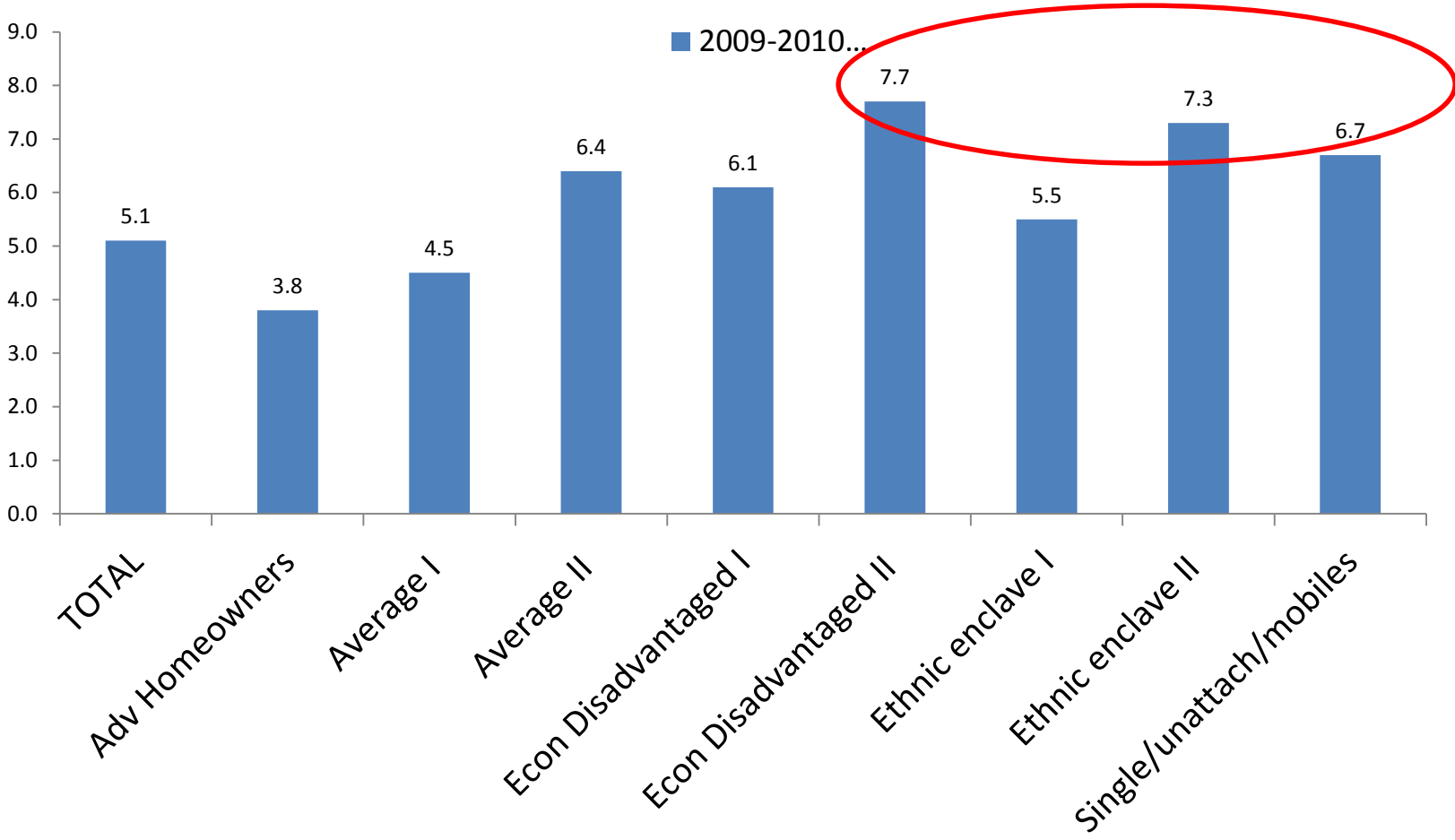
No paid  
advertising

Paid advertising in 2000 and 2010

# ACS March mail-response by segment: Decennial vs non-decennial year

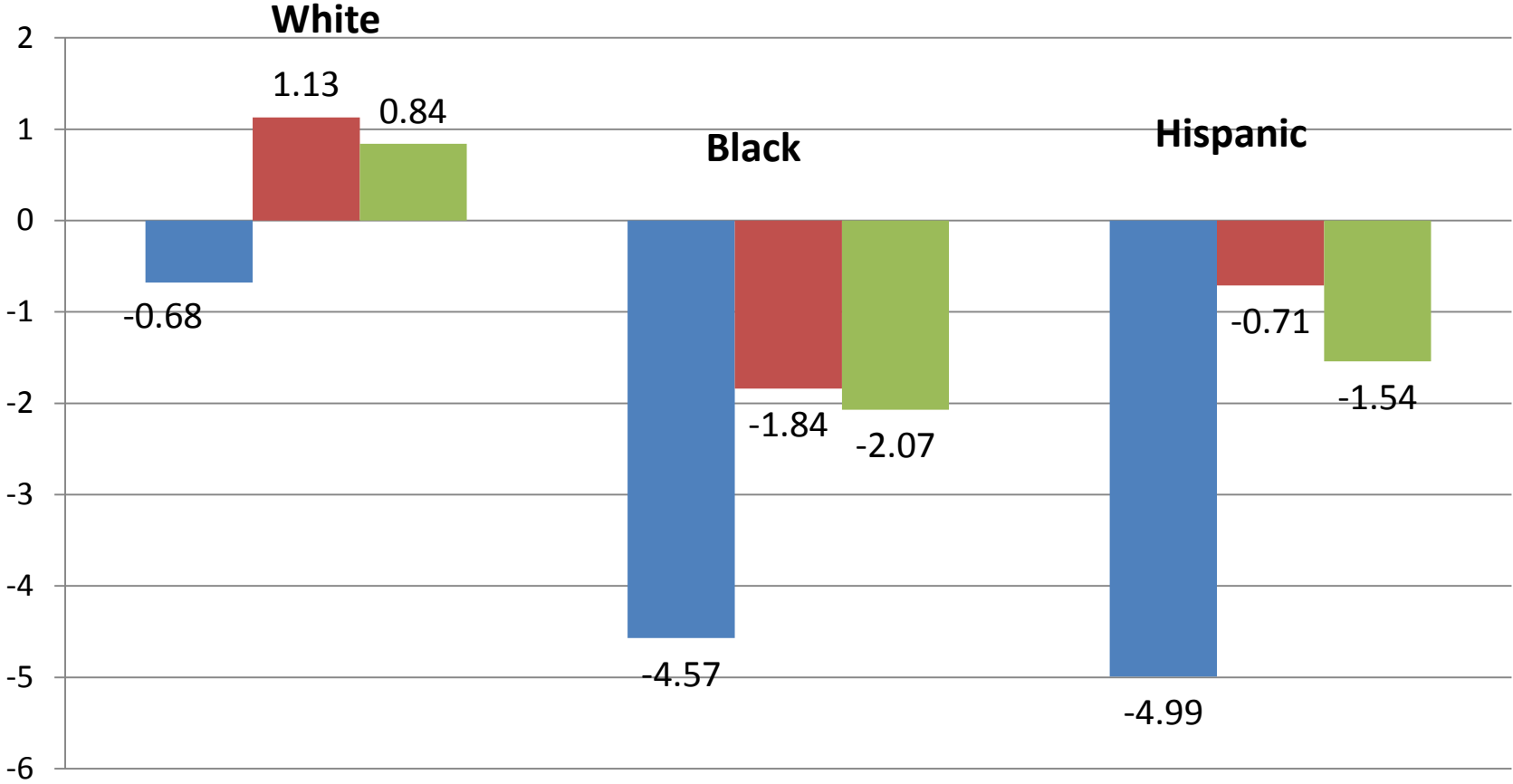


# ACS March mail-response by segment: difference between decennial and non-decennial





# 1990-2010 Census net over/undercounts: by race and ethnicity

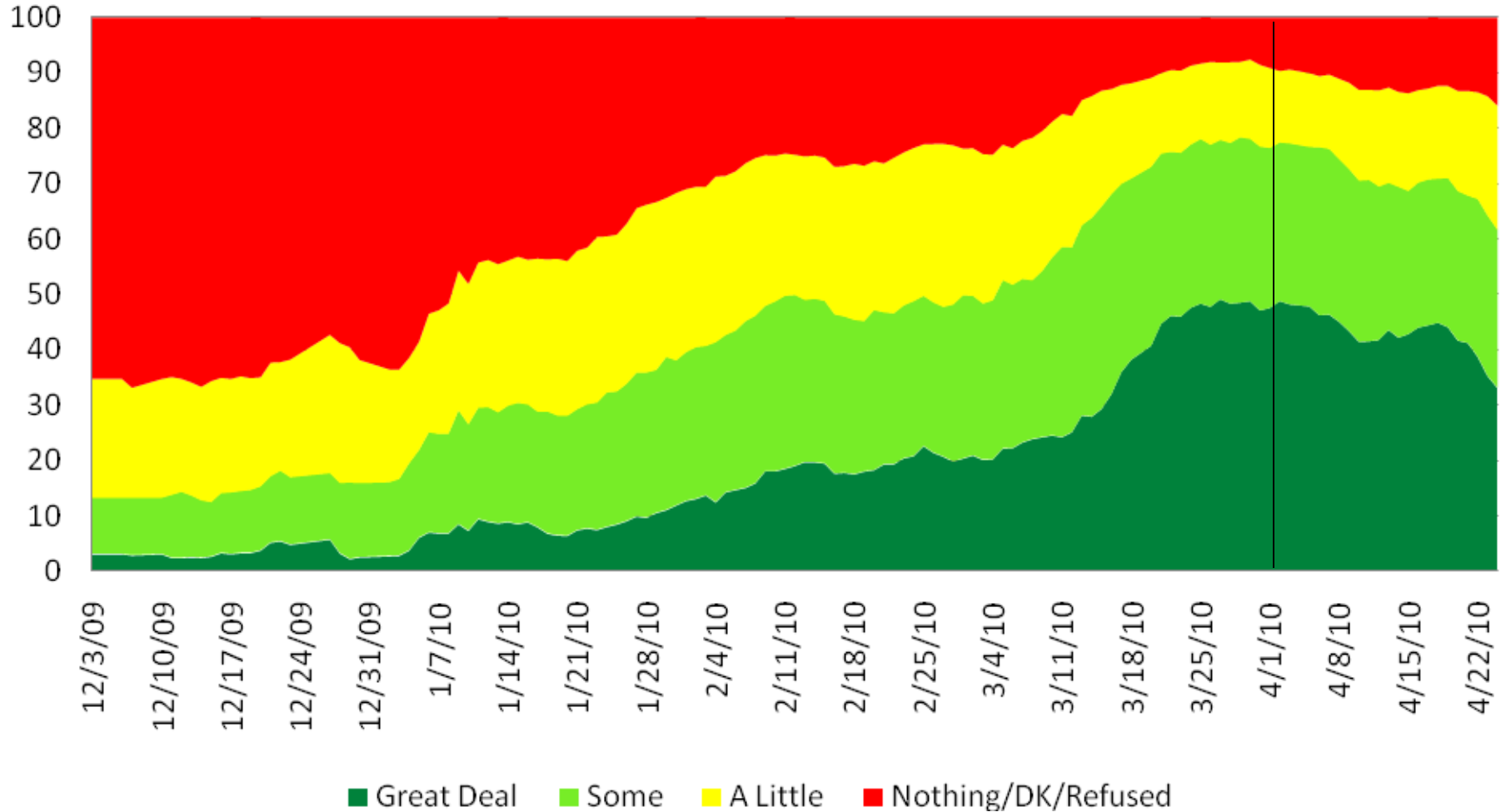


Source: US Census CCM, 2012

■ 1990      ■ 2000      ■ 2010  
No paid advertising      Paid advertising campaigns in 2000 and 2010

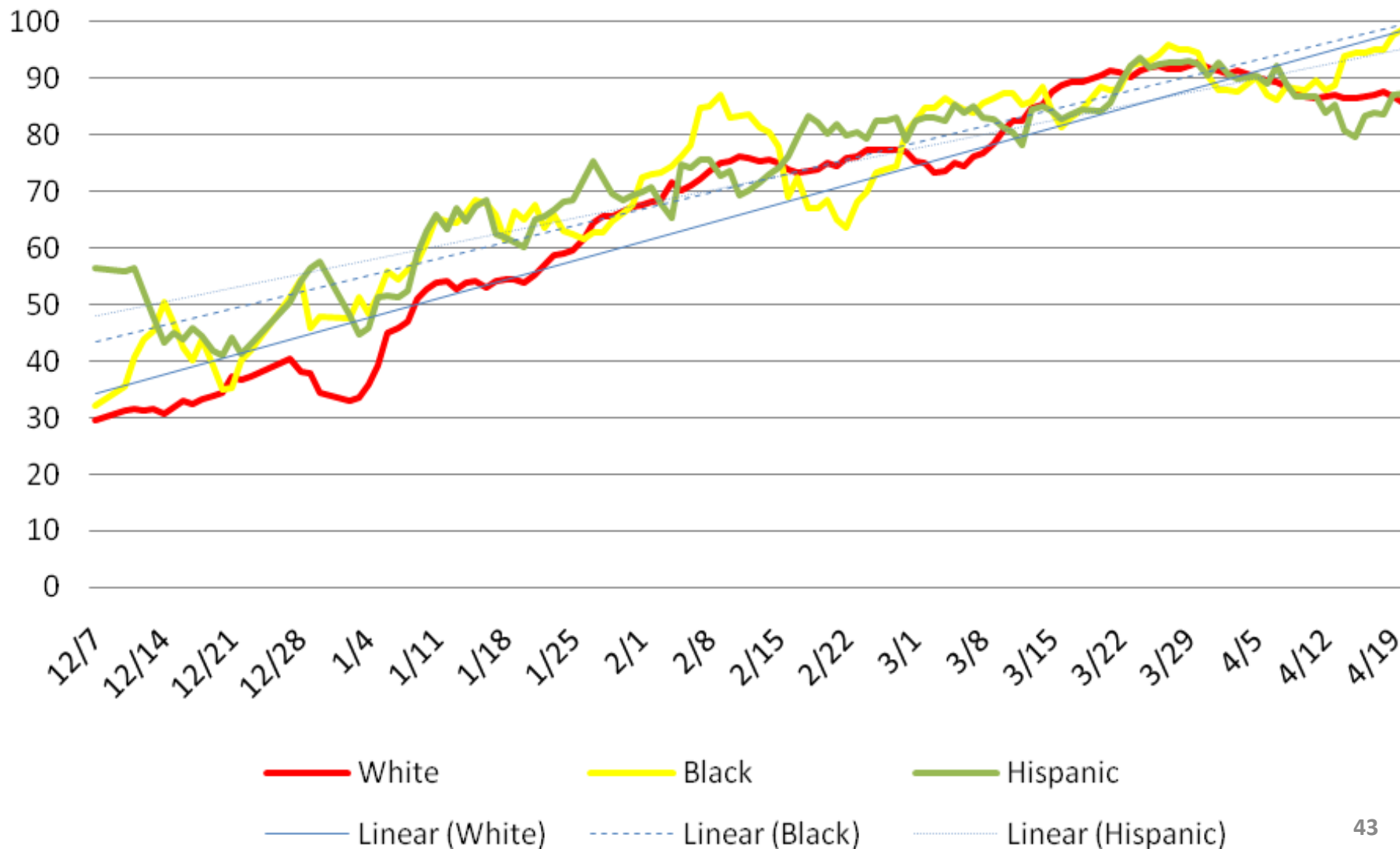
“How much have you seen or heard recently – within the last week or so – about the 2010 Census?”

Rolling Week %

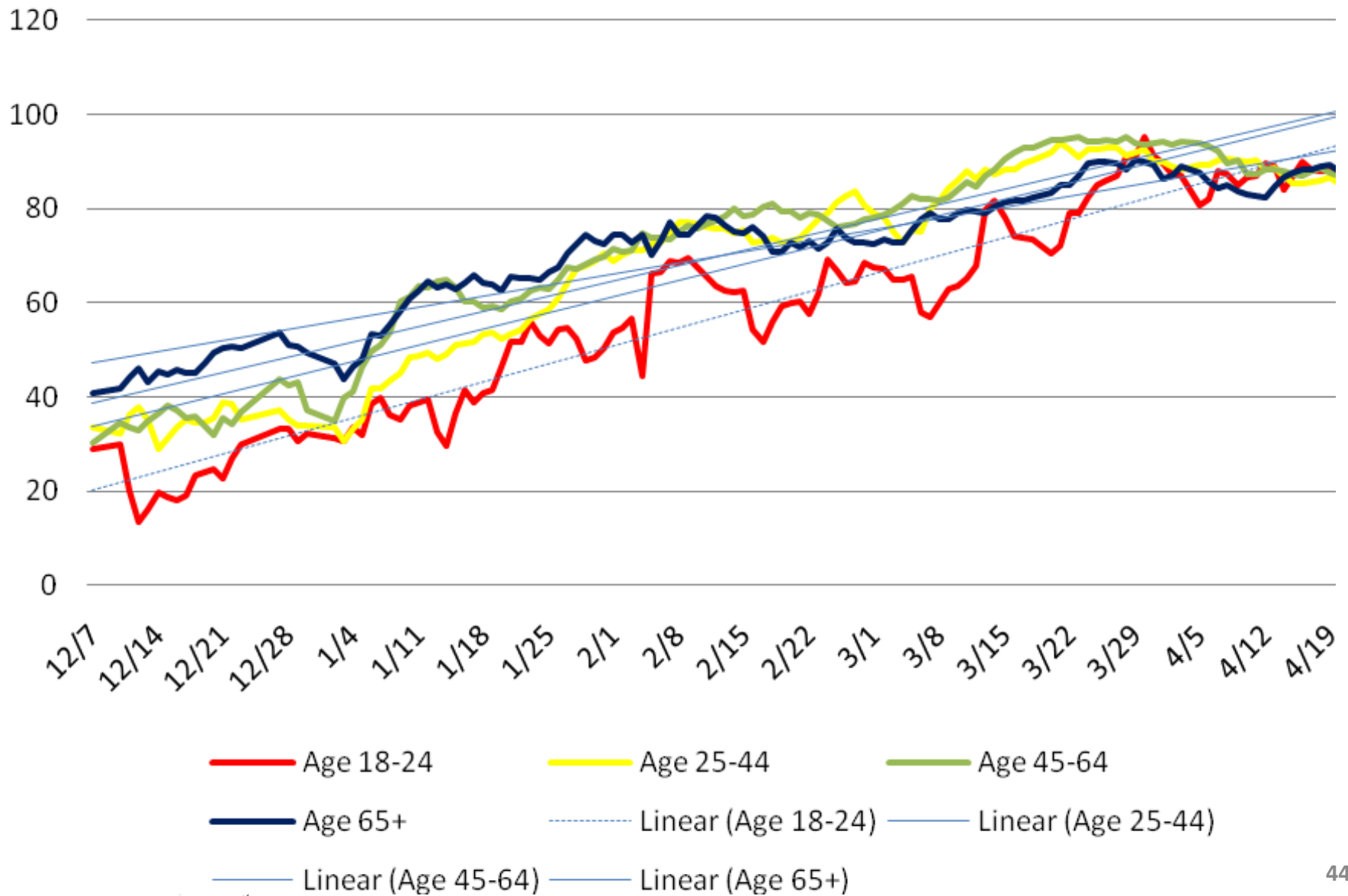


"How much have you seen or heard recently-within the last week or so-  
about the 2010 Census?"

Response: Heard great deal/some/a little  
Rolling Week %



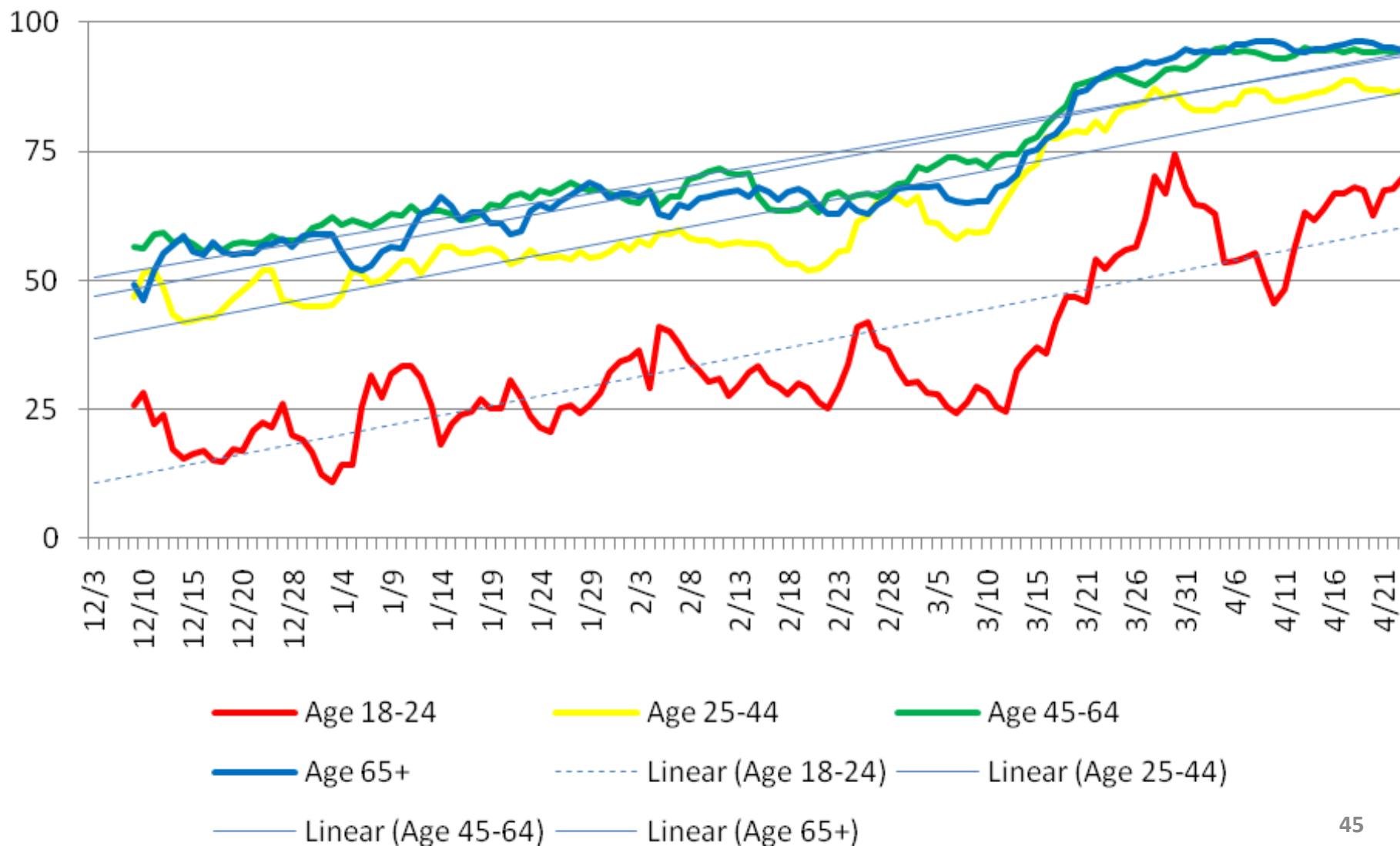
"How much have you seen or heard recently-within the last week or so-about the 2010 Census?" Response: Heard great deal/some/a little  
Rolling Week %



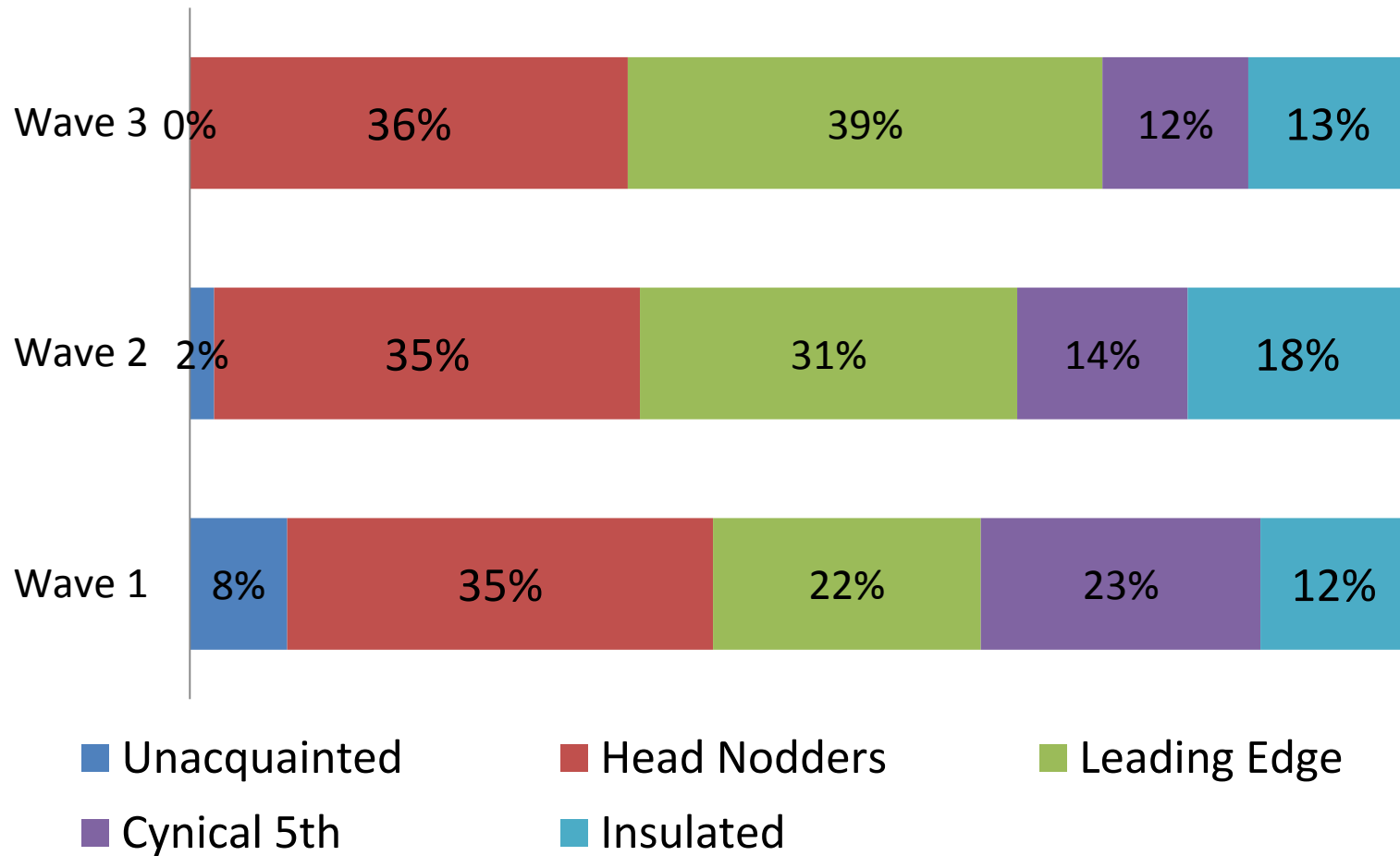
"How likely are you to participate in the 2010 Census? By participate we mean fill out and mail in a Census form"

Response: Definitely Will/Already Mailed Back

Rolling Week %



# 2010 Census ICP Shift in Mindsets (Panel Cases)



# *What did we do right?*

- Segmentation
- Targeted ads
- Real time monitoring
- Rapid response program

# *Where can we improve?*

- Oversaturation
- More precise targeting
- Better alignment of geo segments and mindsets
- Better communication mix for HTS (more “trusted voices” less TV; engage single people)
- Further segmentation of Cynical 5<sup>th</sup>



# *2020 social marketing campaign*

- Social media and *digital advertising* increasing exponentially
- 34% use phone as their primary Internet device\*
  - More likely young adults, non-whites, and low income/education
- Smartphone ownership gap has closed\*
  - Whites, 66%; Black 68% and Hispanics: 64%
- *Mobile devices* as new contact and response method
  - Text/SMS
  - Digital advertisement

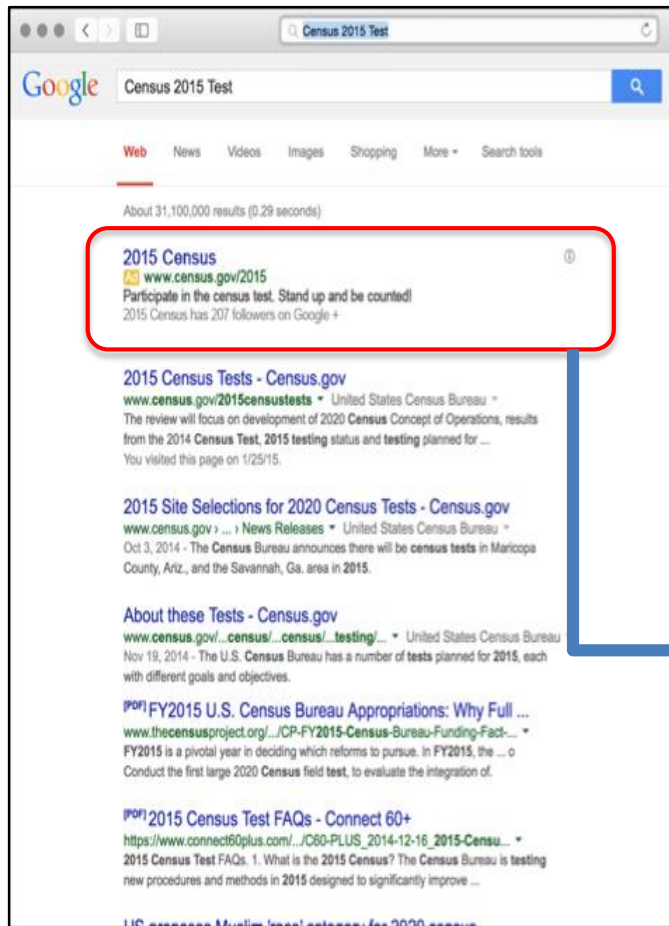
# 2015 Digital Ad Test

- Savannah designated market area (DMA)
- Selection of households received mailing materials
- DMA-wide communications campaign to simulate census environment:
  - Earned Media
  - Partnership outreach
  - Television and radio advertising
  - Print and out-of-home advertising
  - Digital advertising and social media

*First test allowing response without a mailed material and direct digital response*

# Digital ads split into 3 categories...

## 1 Paid Search Ads



2015 Census

Ad [www.census.gov/2015](http://www.census.gov/2015)

Participate in the census test. Stand up and be counted!

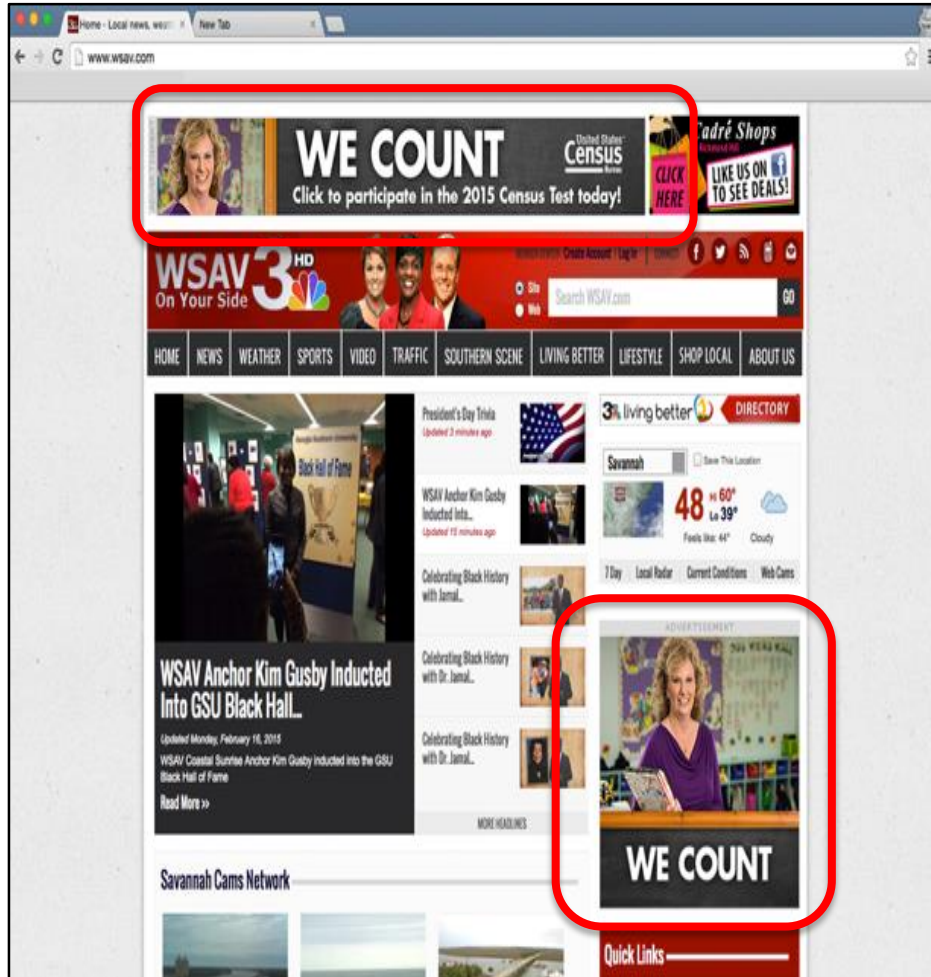
2015 Census has 207 followers on Google +

2

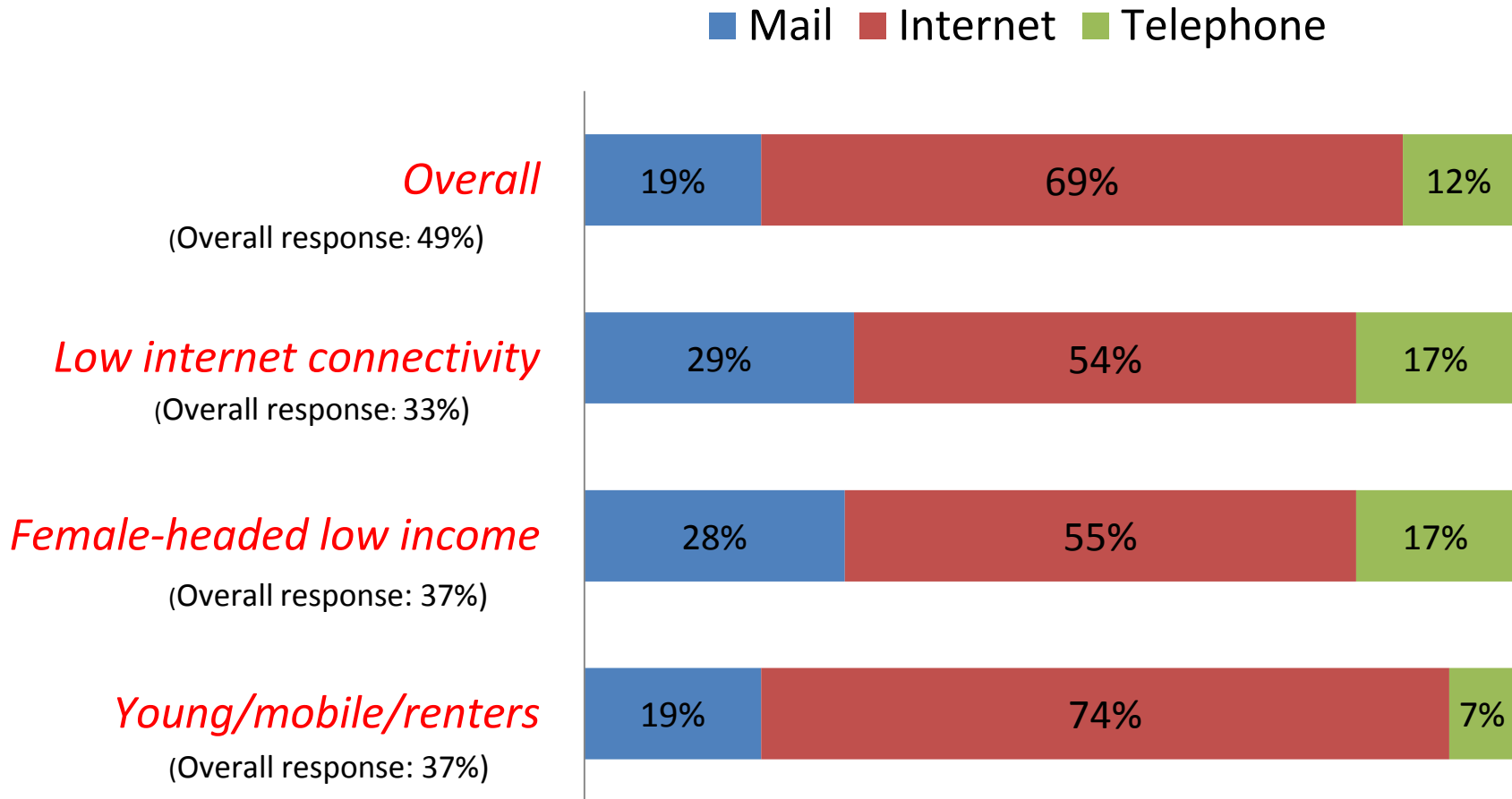
## Social media in-feed



### 3 Display ads

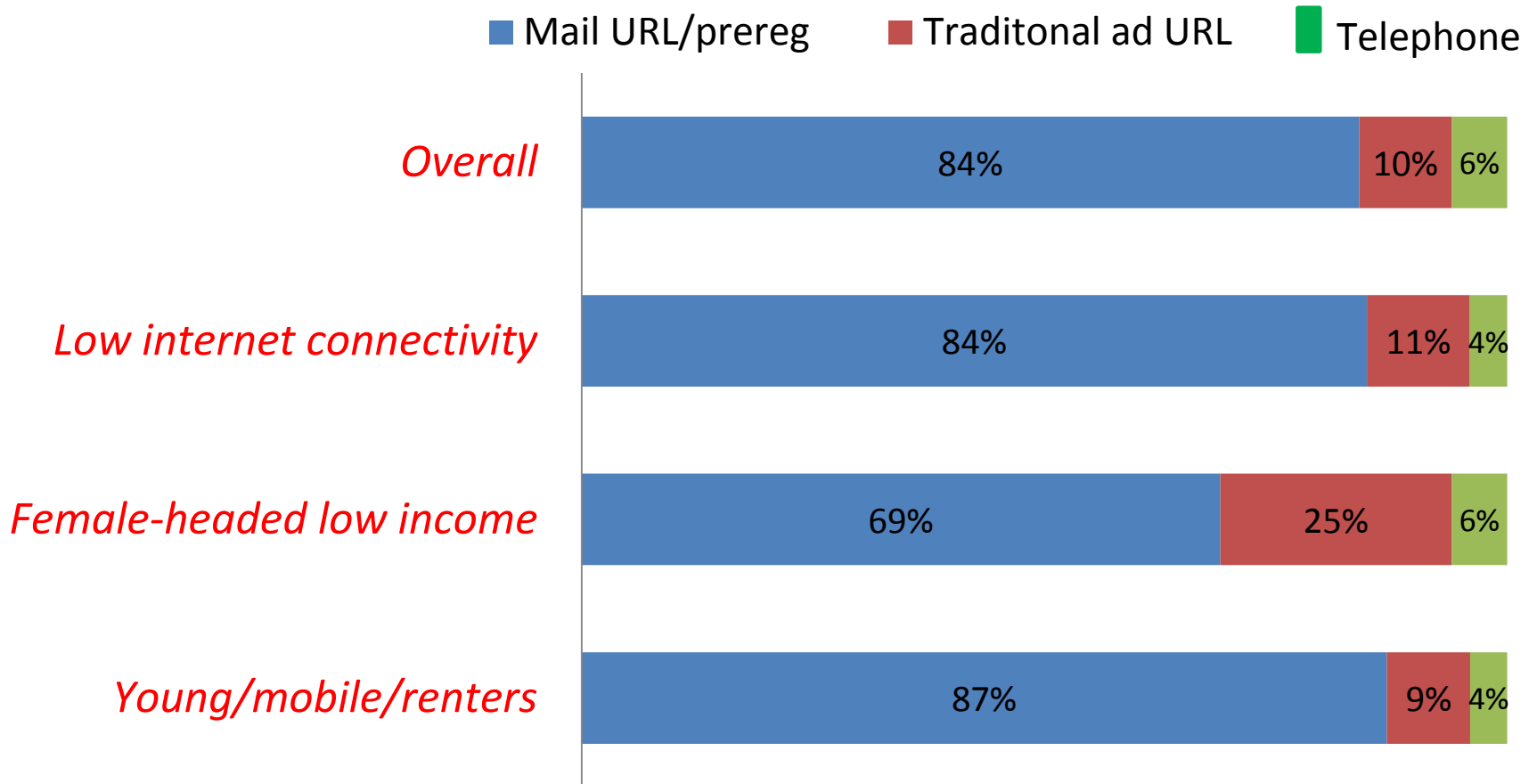


# 2015 Digital Ad Test: response mode by HTS segments: **Mailout** households



Source: Virgile and Bates, 2016

# Source of online response among HTS segments: **Mailout** households



Source: Virgile and Bates, 2016

# *Tagline and awareness campaign to increase self-response in continuing surveys*

- Content: overall **Census Bureau branding** and messaging to increase awareness
- Surveys not mentioned by name and **no immediate call to action** (i.e., click here to complete your survey)
- Deliver ads **to households in sample**



# Video pre-roll example

<https://www.youtube.com/playlist?list=PLhLB6X1ybzTSei5-DCgvsPsonAb7MCUu5>

# *New HTS tools for 2020?*

## THE LOW RESPONSE SCORE (LRS) A METRIC TO LOCATE, PREDICT, AND MANAGE HARD-TO-SURVEY POPULATIONS

---

CHANDRA ERDMAN

NANCY BATES

**Abstract** In 2012, the US Census Bureau posed a challenge under the America COMPETES Act, an act designed to improve the competitiveness of the United States by investing in innovation through research and development. The Census Bureau contracted Kaggle.com to host and manage a worldwide competition to develop the best statistical model to predict 2010 Census mail return rates. The Census Bureau provided competitors with a block group-level database consisting of housing, demographic, and socioeconomic variables derived from the 2010 Census, five-year American Community Survey estimates, and 2010 Census operational data. The Census Bureau then challenged teams to use these data (and other publicly available data) to construct the models. One goal of the challenge was to leverage winning models as inputs to a new model-based hard-to-count (HTC) score, a metric to stratify and target geographic areas according to propensity to self-respond in sample surveys and censuses. All contest winners employed data-mining and machine-learning techniques to predict mail-return rates. This made the models relatively hard to interpret (when compared with the Census Bureau's original HTC score) and impossible to directly translate to a new HTC score. Nonetheless, the winning models contained insights toward building a new model-based score using variables from the database. This paper describes the original algorithm-based HTC score, insights gained from the Census Return Rate Challenge, and the model underlying a new HTC score.

# *New metric to locate HTS: Low Response Score (LRS) model*

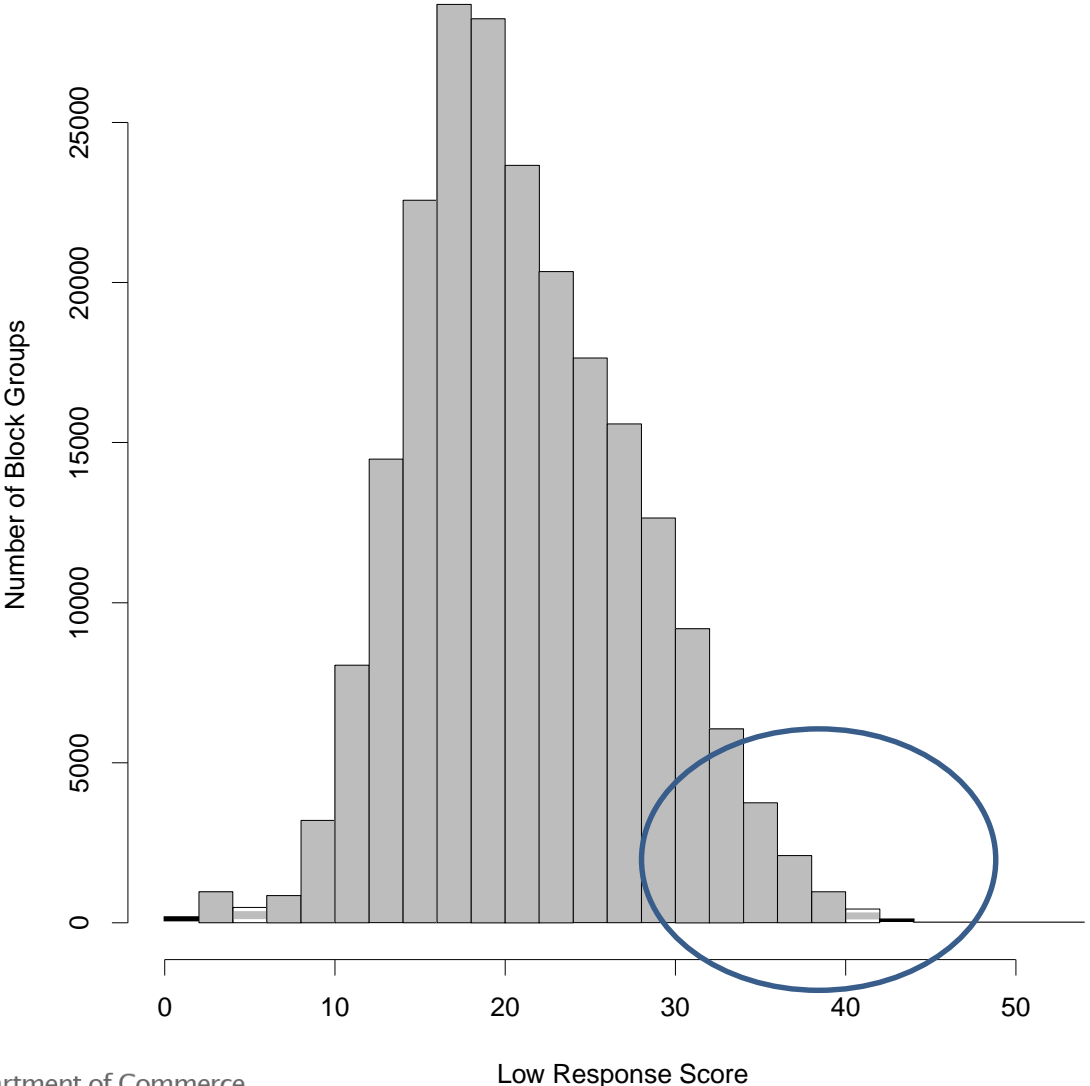
	<b>Coef</b>	<b>Sig</b>		<b>Coef</b>	<b>Sig</b>
(Intercept)	10.29	***	Renter occupied units	1.08	***
Ages 18-24	0.64	***	Female head, no husband	0.58	***
Non-Hispanic White	-0.77	***	Ages 65+	-1.21	***
Related child <6	0.46	***	Males	0.09	***
Married family households	-0.12	***	Ages 25-44	-0.06	
Vacant 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	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	***
Black	-0.04	**	Single person households	-0.24	***
Not high school grad	-0.06	***	Median house value	0.71	***

Sig: \*\*\*  $p < .001$ ; \*\*  $.001 \leq p < .01$ ; \*  $.01 \leq p < .05$

R-squared: 56.10%,  $n = 217,417$

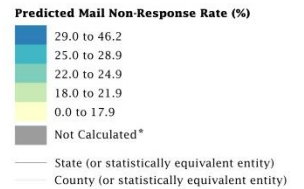
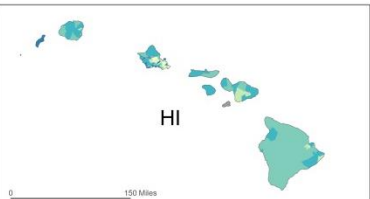
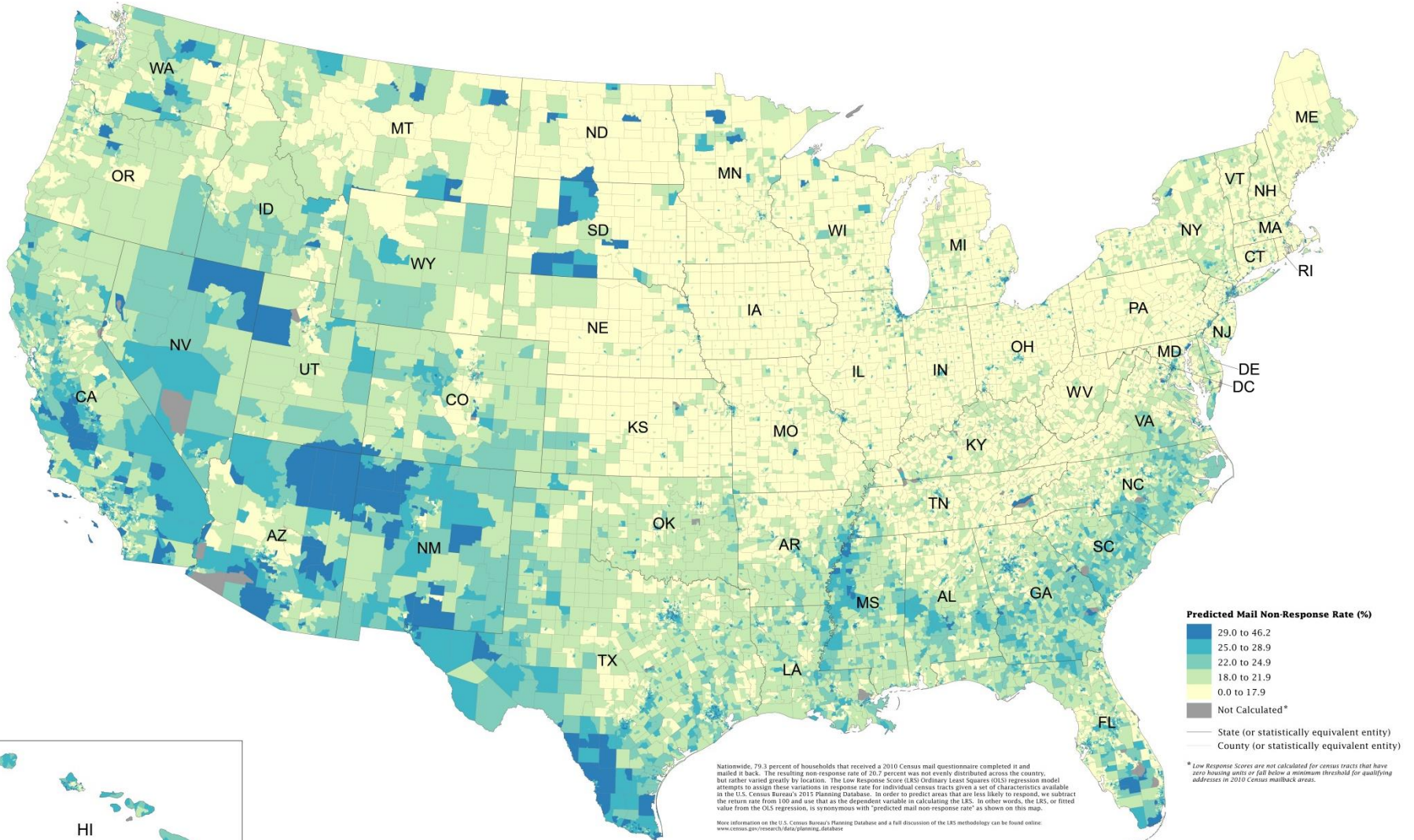
Source: Erdman and Bates, in press

# Distribution of the LRS



# Identifying Hard-to-Survey Populations

## Low Response Score (LRS) by Census Tract



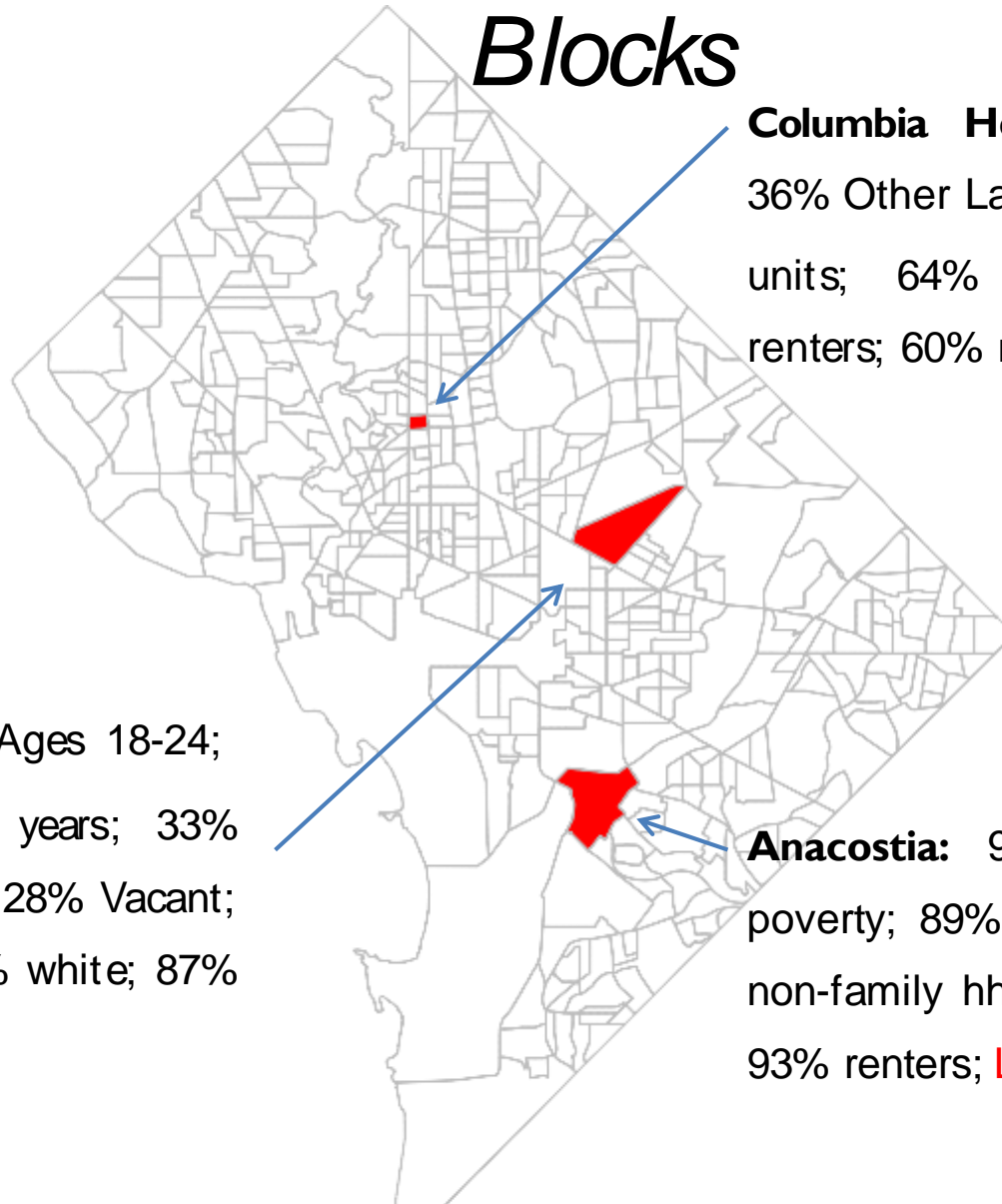
Nationwide, 79.3 percent of households that received a 2010 Census mail questionnaire completed it and mailed it back. The resulting non-response rate of 20.7 percent was not evenly distributed across the country, but rather varied greatly by location. The Low Response Score (LRS) Ordinary Least Squares (OLS) regression model attempts to assign these variations in response rate for individual census tracts given a set of characteristics available in the U.S. Census Bureau's 2010 Planning Database. In order to predict areas that are less likely to respond, we subtract the return rate from 100 and use that as the dependent variable in calculating the LRS. In other words, the LRS, or fitted value from the OLS regression, is synonymous with "predicted mail non-response rate" as shown on this map.

More information on the U.S. Census Bureau's Planning Database and a full discussion of the LRS methodology can be found online: [www.census.gov/research/data/planning-database](http://www.census.gov/research/data/planning-database)

\* Low Response Scores are not calculated for census tracts that have zero housing units or fall below a minimum threshold for qualifying addresses in 2010 Census mailback areas.

# *LRS/PDB Example: Three HTS DC*

## *Blocks*



**Columbia Heights:** 43% Hispanic; 36% Other Language; 92% 10+ multi-units; 64% non-family hhds; 85% renters; 60% moved 5 years ; **LRS=32**

**Trinidad:** 37% Ages 18-24; 59% Moved 5 years; 33% Below poverty; 28% Vacant; 55% Black; 31% white; 87% renters; **LRS=37**

**Anacostia:** 98% Black; 46% below poverty; 89% single unit homes; 15% non-family hhds; 21% moved 5 years; 93% renters; **LRS=38**

# *LRS and PDB publicly available*

- Google “Census Planning Database”
- LRS on both block-group and tract level files
- Available in CSV and API format

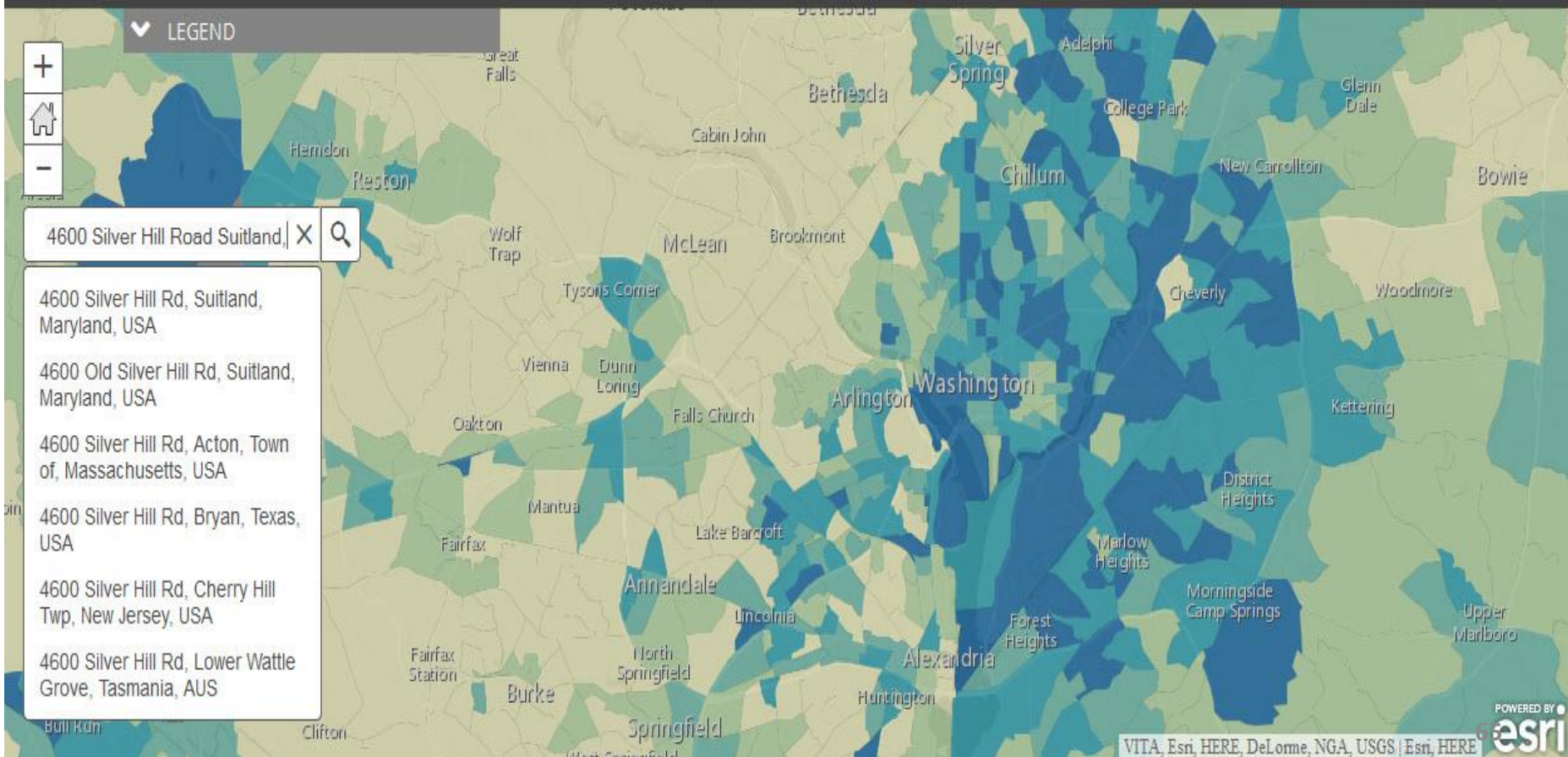


# App for mapping HTS areas

## Low Response Score (LRS) by 2010 Census Tract

Source: U.S. Census Bureau 2015 Planning Database [f](#) [t](#) [e](#)

The LRS is simply the Predicted Mail Non-Response Rate (%) as calculated in the Planning Database. More information on the U.S. Census Bureau's Planning Database and a full discussion of the LRS methodology can be found by clicking the "Source" note to the right.



# App for mapping HTS areas

## Low Response Score (LRS) by 2010 Census Tract

Source: U.S. Census Bureau 2015 Planning Database



The LRS is simply the Predicted Mail Non-Response Rate (%) as calculated in the Planning Database. More information on the U.S. Census Bureau's Planning Database and a full discussion of the LRS methodology can be found by clicking the "Source" note to the right.



### LEGEND

#### Predicted Mail Non-Response Rate (%)

#### LRS

- 29.0 - 46.2
- 25.0 - 28.9
- 22.0 - 24.9
- 18.0 - 21.9
- 0.0 - 17.9
- Not Calculated\* (-1)

Predicted Mail Non-Response Rate (%): 30.0

- 9.5% poverty
- 93.3 % Black
- 3.7% Hispanic
- 60.3% Renters
- 1.6% Limit English
- 11.0% < HS
- 32.5% Female head



Marlow Heights

# LRS limitations

- Only considers mail self-response
- Some tracts have small Ns in mailback universe
- January 2013 ACS began collecting internet access Q; also offered internet as a mode
- First order of business is to update LRS

# *Plans for 2020 campaign*

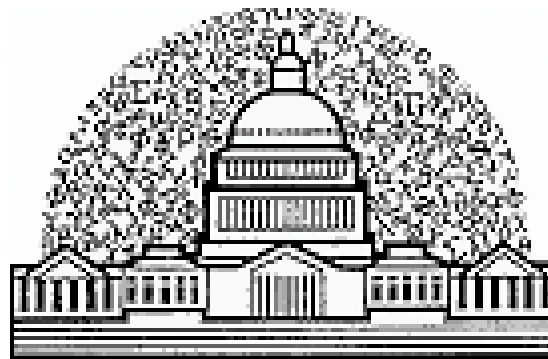
- 2020 Communications Contract awarded to **Young and Rubicam (Y&R)**
  - Multicultural advertising partner subcontractors:
    - *BRAVO, Carol H Williams, Culture ONE World, g+g advertising, TDW & Co; Kalaimoku Group*
- Plan to conduct **household level models**:
  - Propensity to self-respond
  - Preferred mode of response
  - Preferred mode of contact (media vehicle)

# Parting thoughts...

- Post-2010 Census attitudes
- Insulated & Cynical Fifth still aligned
- Trust in government/attitudes became important
- Emerging “**suspicious**” mindset
  - Census will personally harm
  - Misinformed about Census uses
  - Low intent to participate

[nancy.a.bates@census.gov](mailto:nancy.a.bates@census.gov)

Thanks to the Hansen Nomination Committee,  
NASS, WSS and Westat



Washington  
Statistical Society



# Psychographics

Tapestry <sup>®</sup> segment	Female headed low income/education
<u>Modest Income Homes</u>	35%
<u>Rural Bypass</u>	20%

LifeMode Group: Hometown

# Modest Income Homes

Households: 1,632,000

Average Household Size: 2.56

Median Age: 36.1

Median Household Income: \$22,000

## WHO ARE WE?

Families in this urban segment may be nontraditional; however, their religious faith and family values guide their modest lifestyles. Many residents are primary caregivers to their elderly family members. Jobs are not always easy to come by, but wages and salary income are still the main sources of income for most households. Reliance on Social Security and public assistance income is necessary to support single-parent and multigenerational families. High poverty rates in this market make it difficult to make ends meet. Nonetheless, rents are relatively low (Index 73), public transportation is available, and Medicaid can assist families in need.

## OUR NEIGHBORHOOD

- Households are single person or single parent (usually female householders). Multigenerational families are also present.
- Homes are predominantly single family; values reflect the age of the housing, built more than 60 years ago.
- Over half of the homes are renter occupied; average rent is lower than the US average.
- Most households have no vehicle or one car, with a high dependence on public transportation.

## SOCIOECONOMIC TRAITS

- Almost a quarter of adults aged 25 or more have no high school diploma.
- Labor force participation is only 50%, with unemployment at more than double the US rate.
- Income is less than half of the US median income; one in three households are in poverty, dependent on Social Security, public assistance, and Supplemental Security Income.
- Consumers in this market consider traditional gender roles and religious faith very important.
- This market lives for today, choosing to save only for a specific purpose.
- They favor TV as their media of choice and will purchase a product with a celebrity endorsement.





LifeMode Group: Hometown

# Modest Income Homes

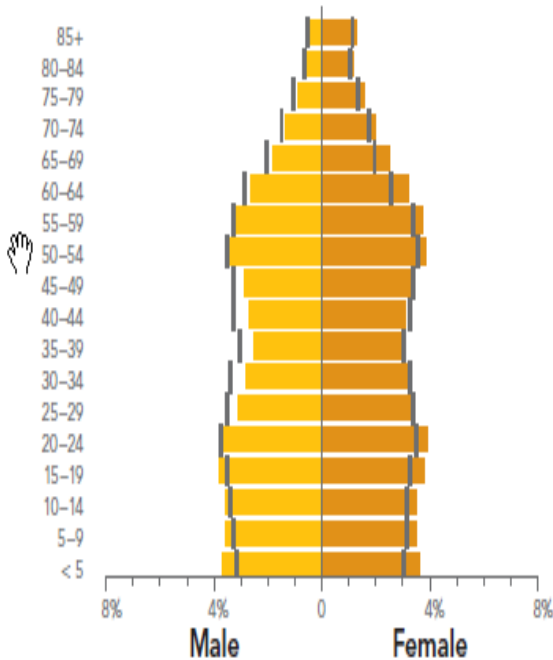


TAPESTRY SEGMENTATION  
esri.com/tapestry

## AGE BY SEX (Esri data)

Median Age: **36.1** US: 37.6

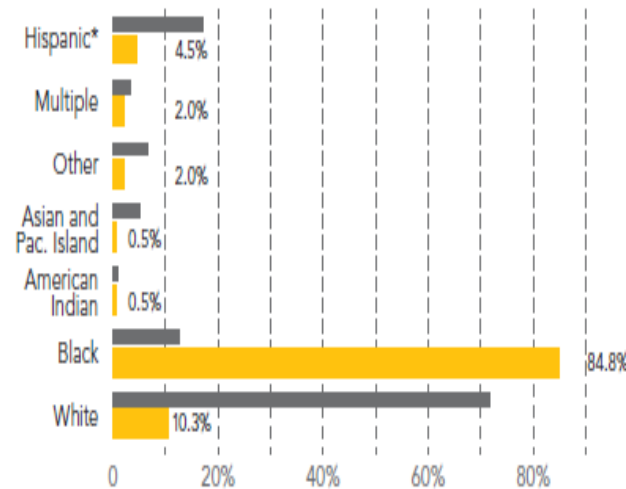
| Indicates US



## RACE AND ETHNICITY (Esri data)

The Diversity Index summarizes racial and ethnic diversity. The index shows the likelihood that two persons, chosen at random from the same area, belong to different race or ethnic groups. The index ranges from 0 (no diversity) to 100 (complete diversity).

Diversity Index: **33.3** US: 62.1

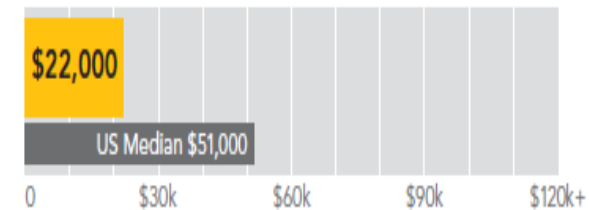


\* Hispanic can be of any race. US Average

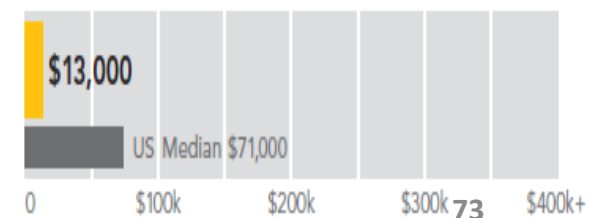
## INCOME AND NET WORTH

Net worth measures total household assets (homes, vehicles, investments, etc.) less any debts, secured (e.g., mortgages) or unsecured (credit cards). Household income and net worth are estimated by Esri.

### Median Household Income



### Median Net Worth





# Modest Income Homes



## MARKET PROFILE (Consumer preferences are estimated from data by GfK MRI)

- To make ends meet, consumers shop at warehouse clubs and low-cost retailers.
- Unlikely to own a credit card, they pay their bills in person.
- This market supports multigenerational families; they are often primary caregivers for elderly family members. On average, *Modest Income Homes* residents have a higher reliance on Medicaid.
- For entertainment, they listen to gospel and R&B music and prefer to watch BET.
- The recreational activity of choice for residents is basketball.

## HOUSING

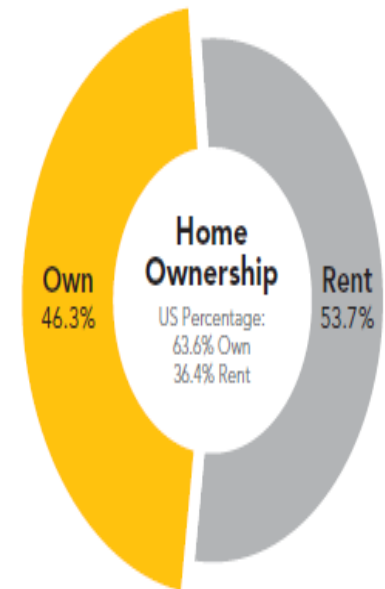
Median home value is displayed for markets that are primarily owner occupied; average rent is shown for renter-occupied markets. Tenure and home value are estimated by Esri. Housing type and average rent are from the Census Bureau's American Community Survey.



Typical Housing:  
Single Family

Average Rent:  
\$720

US Average: \$990



# Self Response by Mode – Percent of Total Submissions

Self Response Mode	All Submissions	
	Count	% of Submissions
<b>Total Submissions</b>	<b>70,208</b>	<b>100.0%</b>
<b>Mode: Internet</b>	<b>56,145</b>	<b>80.0%</b>
Blank (largely mailout URL)	27,171	38.7%
Census Vanity URL	19,948	28.4%
Digital Ad Initiated	7,704	11.0%
Partners/Events	287	0.4%
GovDelivery (email)	230	0.3%
Postcard URL	764	1.1%
QR Code (from print ads)	19	0.0%
Out-of-Home SMS	12	0.0%
Organic Social Media	5	0.0%
Unidentified	5	0.0%
<b>Mode: Paper, mail-back</b>	<b>8,461</b>	<b>12.1%</b>
<b>Mode: Telephone, in-bound</b>	<b>5,602</b>	<b>8.0%</b>

**40.1%** directly attributable to advertising and communications campaign