

# Using the Planning Database to Improve the Survey Process

Kathleen M Kephart

US Census Bureau  
AAPOR Conference Austin, TX

May 14, 2016

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau

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

FUTURE ON  
Activating Change

1

## Overview

- A “greatest hits” of ACS 5 year estimates and 2010 Census variables
- Pulls together publicly available estimates in one convenient file
- Available at two levels of geography: Tract and Block Group

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

FUTURE ON  
Activating Change

2

## Contents of the 2015 PDB

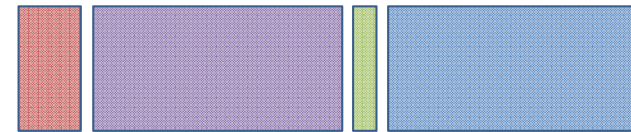
- Both 5-year ACS estimates and 2010 census data
- Over 200 variables, including, but not limited to:
  - Population: gender, age, education, poverty, health insurance coverage, etc.
  - Household: language, relationship, income, etc.
  - Housing unit: tenure, number of units, etc.
  - Census operational: mailout/mailback, bilingual, etc.

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

FUTURE ON  
Activating Change

3

## 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\_09\_13, Males\_ACSMOE\_09\_13

### 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\_09\_13, pct\_Males\_ACSMOE\_09\_13

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

FUTURE ON  
Activating Change

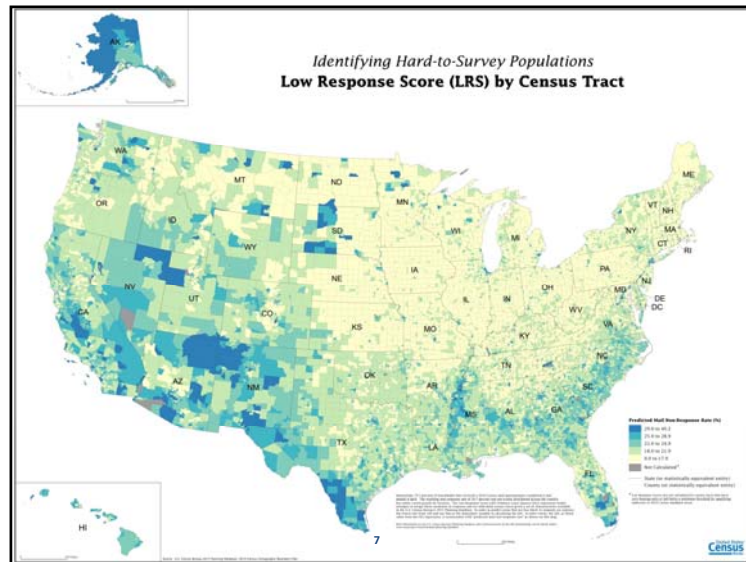
4

## Potential Uses for the PDB in the Survey Process

- Identify areas with a predicted high mail non-response
- Stratify a sample on key variables
- Identify areas that require Non-English materials
- Identify potential non-response bias

## The Low Response Score (Erdman and Bates, 2014)

- Similar to the 2000 Census Hard-to-Count Score
- A model based predicted mail response score
- Uses key predictors of mail non-response including
  - Household composition (Female householders, etc)
  - Tenure (percent of Renters)
  - Other key predictors and more information can be found in (Erdman and Bates, Forthcoming 2016)
- Can be used to target areas with potential high non-response



## 2016 Census Test Harris County Texas Demographics

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

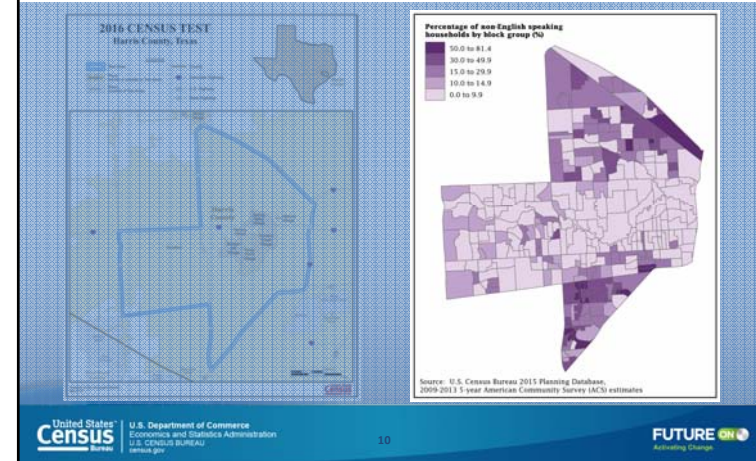
	Houston*	United States*
Households where no one over 14 speaks English "very well"	14.8%	4.6%
Population 18-24 years old	9.4%	10.0%
Renter Occupied Units	60.9%	35.1%
Population 25 and over, with less than a HS diploma	19.1%	13.9%

Source: 2015 PDB Block Group File [http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
\*ACS 5 year 2009-2013 Estimate factfinder.census.gov

## Limited English Proficiency

- Identify areas that need additional language materials
- Flag block groups or tracts that have a high percentage of housing units where no one over the age of 14 speaks English “very well”
- Identify what language is spoken in these block groups or tracts

## Where are non-English speaking households located?



## Limited English Proficiency Block Groups 2016 Census Harris TX Test Site

Block Group	No one speaks English “very well”	Spanish	Asian/Pacific Islander	Other
4327012	81.4% (14.3)	81.4% (14.0)	0% (2.1)	0% (2.1)
4330012	77.2% (13.4)	73.4% (13.5)	3.8% (4.1)	0% (2.3)
4327011	72.5% (11.1)	72.5% (10.9)	0% (1.6)	0% (1.6)
4335012	69.3% (10.9)	66.1% (10.7)	0% (1.7)	3.2% (4.8)
5214001	69.3% (21.1)	69.3% (20.6)	0% (3.7)	0% (3.7)

Source: 2015 PDB Block Group File [http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
 \*ACS 5 year 2009-2013 Estimate factfinder.census.gov

## Identifying Potential Non-response Bias

- If you know what tracts or block groups have high non-response in your survey you can estimate how non-respondents may differ on key demographics from respondents

	Respondent Characteristics	Sample Tract with high Non-response*
Percent living in poverty	5.3% (1.2)	59.6% (9.3)
Population 25 and over, with less than a HS diploma	19.1% (2.3)	28.5% (10.6)
Renter Occupied Units	35.1% (3.2)	86.4% (5.9)

Source: 2015 PDB Tract File [http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
 ACS 5 year 2009-2013

## Downloading Data to Build a Custom PDB

- ACS 5 year Estimates and 2010 Census Data are available at other geographies including
  - Zip Code Tabulation Areas
  - Congressional Districts
  - Counties
  - Other geographies as well



## What are Zip Code Tabulation Areas (ZCTA)?

- Address's USPS Zip Code and ZCTA can be different!
- Generalized areal representations of United States Postal Service Zip Codes
- Only respect Census block boundaries, cross all other geographic boundaries
- Most common Zip Code within a block is assigned to all addresses in the block
- If all Zip Codes are equal, assign based on neighboring block's ZCTA

## Example: ZCTA 60565

- 11 tracts that cross 2 counties are all or partially contained within this Zip Code Tabulation Area

Zipcode	County Name	TractID	Percent of pop in tract	Percent of land area in tract
60565	DuPage	846203	8.9%	7.4%
60565	Will	880118	8.9%	10.9%

Source: [https://www.census.gov/geo/maps-data/data/zcta\\_rel\\_download.html](https://www.census.gov/geo/maps-data/data/zcta_rel_download.html)

## How to Download (1)



## How to Download (2)

**Download Center - A step-by-step guide to downloading data**

1 Start 2 Dataset 3 Geographies 4 Search Results

Select geographies to add to Your Selections. Click Next

The download center allows you to select geographic areas, such as all counties in a state. To select from all available geographies, use Advanced Search.

Select a geographic type  
5-Digit ZIP Code Tabulation Area - 900

Select a state  
-- select a state --

Select one or more geographic areas and click Add to Your Selections  
Add 5-Digit ZIP Code Tabulation Areas within United States and Puerto Rico

ADD TO YOUR SELECTIONS

**Your Selections**

Search using:  
Dataset: 2013 ACS 5-year estimates

Tables matching your selections: 2,290

PREVIOUS NEXT CANCEL

United States Census Bureau U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov 17 FUTURE ON Activating Change

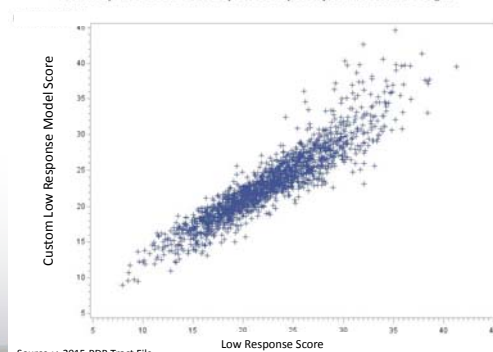
## LRS Vs. Custom Low Response Model

- Created an OLS model similar to Low Response Score on Georgia tract level PDB, used ACS 2009-2013 Estimates as Predictors
- Not the same as official Low Response Score
- Uses just recent ACS estimates

United States Census Bureau U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov 18 FUTURE ON Activating Change

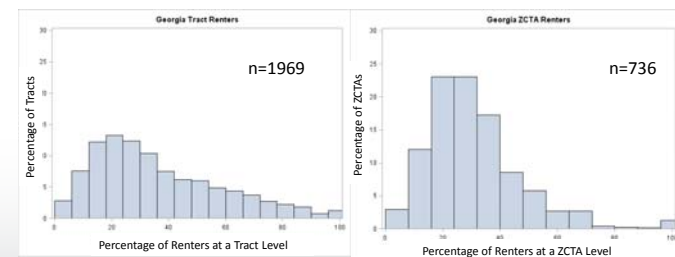
## Low Response Score vs. Custom Low Response Model

Low Response Score VS. Response Propensity for Tracts in Georgia



United States Census Bureau U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov 19 FUTURE ON Activating Change

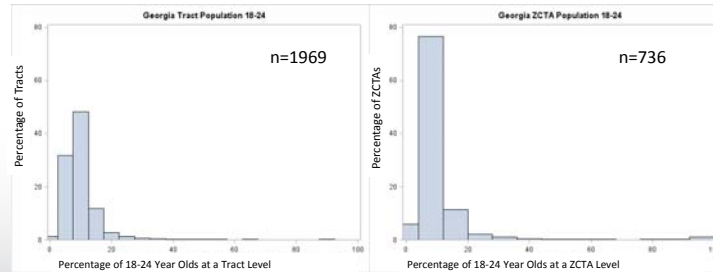
## Percentage of Renters Tract vs ZCTA Level



Source: 2015 PDB Tract File  
[http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
and ZCTA ACS 5 year 2009-2013 Estimate factfinder.census.gov

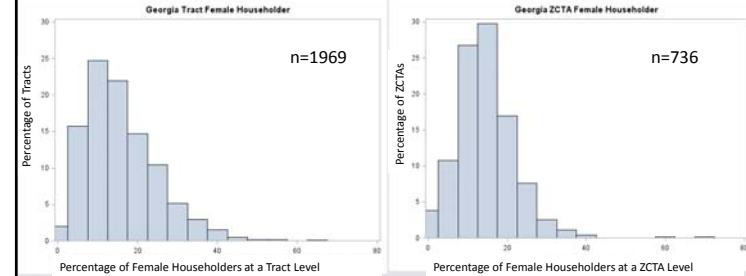
United States Census Bureau U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU census.gov 20 FUTURE ON Activating Change

## Percentage of 18-24 Year Olds Tract vs ZCTA Level



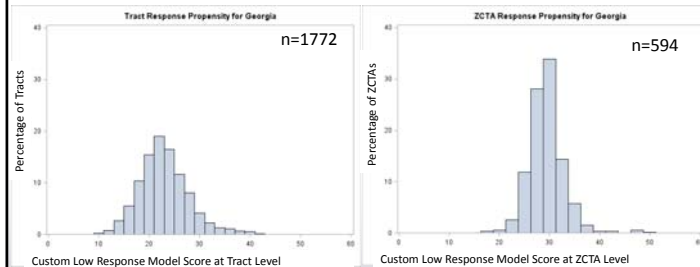
Source: 2015 PDB Tract File  
[http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
 and ZCTA ACS 5 year 2009-2013 Estimate factfinder.census.gov

## Percentage of Female Householders Tract vs ZCTA Level



Source: 2015 PDB Tract File  
[http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
 and ZCTA ACS 5 year 2009-2013 Estimate factfinder.census.gov

## Tract Level Custom LRS Coefficients Applied to ZCTA Level



Source: 2015 PDB Tract File  
[http://www.census.gov/research/data/planning\\_database/2015/](http://www.census.gov/research/data/planning_database/2015/)  
 and ZCTA ACS 5 year 2009-2013 Estimate factfinder.census.gov

## Limitations

- 2010 Census Operational Data (including 2010 Mail Return Rate) is only available at tract and block group level
- Our Custom Low Response model is an approximation of Low Response Score
  - Uses only recent ACS estimates at a tract level to build coefficients
  - Does not perform well when scored on ZCTA level
  - Predicts only mail response, not other modes



## Acknowledgements

- Thank you to Travis Pape and Julia Coombs for creating the code to generate the PDB
- Luke Larsen and Alina Kline for their work on the upcoming 2016 PDB
- Nancy Bates and Barb O'hare for their time and effort to bring the PDB back
- Suzanne McArdle for her work on PDB data visualizations

## How to Access the PDB and Contact Info

- Available on the Census Bureau's Research @ Census page
- Link to the PDB CSV format:  
[http://www.census.gov/research/data/planning\\_database/](http://www.census.gov/research/data/planning_database/)
- API format: [www.census.gov/developers](http://www.census.gov/developers)

### Questions?

Kathleen.m.kephart@census.gov