

# Low Response Score for Use in Survey and Census Planning and Analysis

**Chandra Erdman and Nancy Bates**  
U.S. Census Bureau

**Disclaimer:** The views expressed on statistical issues are those of the authors only.

- 1 The original Hard-to-Count (HTC) Score
- 2 The Census Kaggle Challenge
- 3 The Low Response Score (LRS)

# The Original HTC Score

Bruce et al. (2001); Bruce and Robinson (2003)

- |                            |                                      |
|----------------------------|--------------------------------------|
| ① Renter occupied units    | ⑦ Different housing unit 1 year ago  |
| ② Unmarried                | ⑧ Public assistance                  |
| ③ Vacant units             | ⑨ Unemployed                         |
| ④ Multi-unit structures    | ⑩ Crowded units                      |
| ⑤ Below Poverty            | ⑪ Linguistically isolated households |
| ⑥ Not high school graduate | ⑫ No phone service                   |

# The Census Kaggle Challenge

*"All you need is data and a question. Our data scientists will provide the answer."*

*– Kaggle.com*

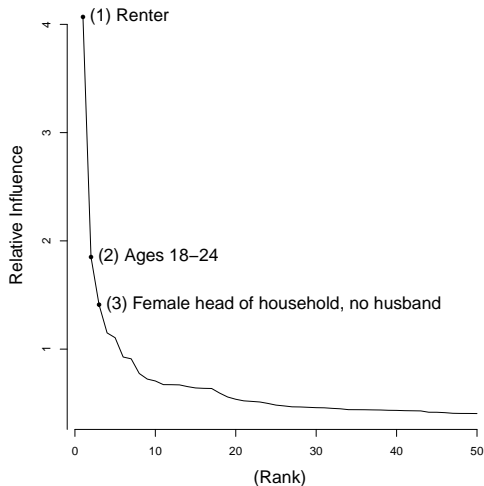
Data: 2012 Block-Group-Level Planning Database (PDB)

Question: Which statistical model best predicts 2010 Census mail return rates?

Product: Updated model-based "Hard-to-Count" Score

# Winning Model Predictors

- When ranked by relative influence, 24/25 top predictors from PDB



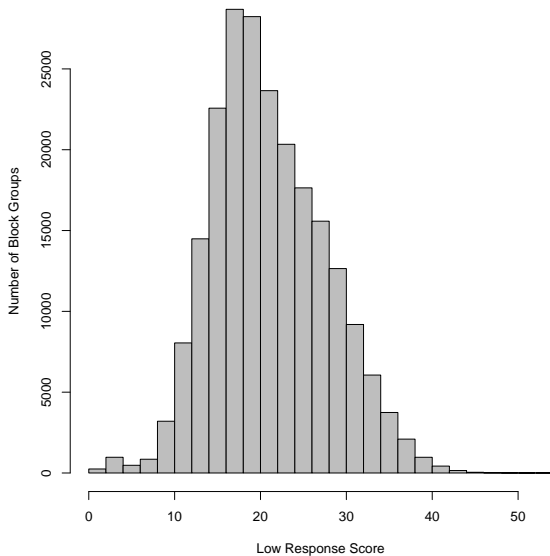
# Low Response Model (Block-Group)

	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	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	***

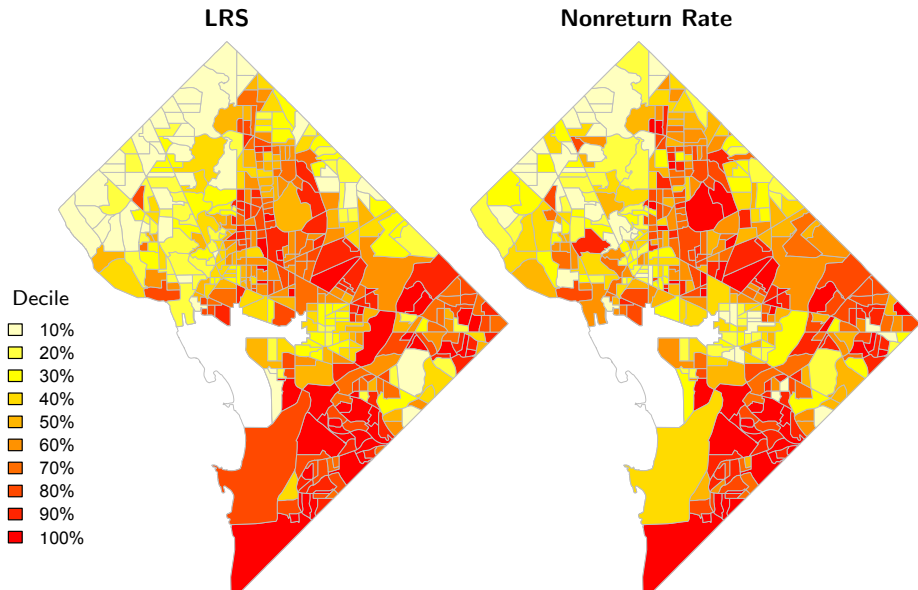
Sig: \*\*\*  $p < .001$ ; \*\*  $.001 \leq p < .01$ ; \*  $.01 \leq p < .05$

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

# Distribution of the LRS

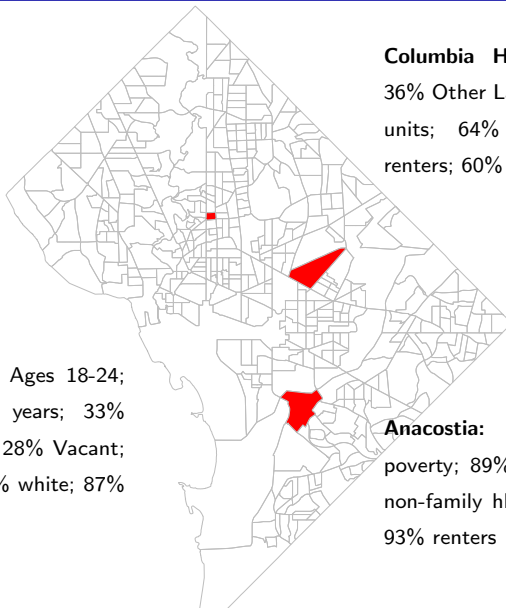


# Deciles of the LRS for Block-Groups in DC





# LRS/PDB Example 1: Three HTC Blocks in DC

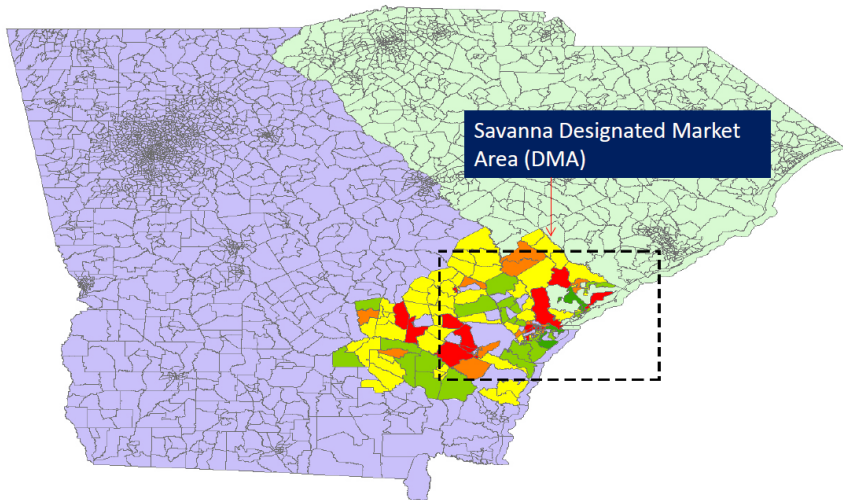


**Columbia Heights:** 43% Hispanic;  
36% Other Language; 92% 10+ multi-  
units; 64% non-family hhds; 85%  
renters; 60% moved 5 years

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

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

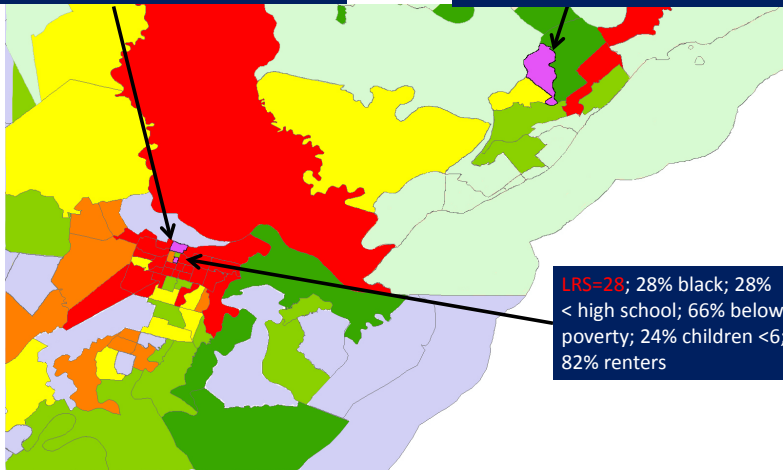
# LRS/PDB Example 2: Targeted Digital Advertising Experiment 2015 Census Site Test



# LRS/PDB Example 2 (Cont.)

Low Response Score=26; 37% 18-24; 34% moved 5 years; 51% single person hhds; 79% not married

Low Response Score=8 ; 47% 65+ ; 84% single units; 96% White; 8% renters



LRS=28; 28% black; 28% < high school; 66% below poverty; 24% children <6; 82% renters

# Considerations

- Independent variable is mail response; 2020 Census will have an Internet response option
- “Single Unattached Mobiles” (Bates and Mulry, 2011)
  - 64.7 percent of American Community Survey self response by Internet (Baumgardner, 2013)
- In January, 2013, ACS began asking about Internet connectivity

# Summary

- Challenge was successful
- Winning model was complex but predictors in rank order of influence proved useful
- Accurate predictions with relatively few predictors
- Simple Low Response Score: OLS predictions
- First score at this level of geography
- Useful for planning and targeted advertising

`julia.coombs@census.gov`

`barbara.c.ohare@census.gov`

`chandra.erdman@census.gov`

`nancy.a.bates@census.gov`