



# Use of Administrative Records to Reduce Burden and Improve Quality

Amy O'Hara  
CNSTAT Workshop on Respondent Burden  
March 8, 2016

# Using administrative records in household surveys

- Reduce burden
  - American Housing Survey (AHS)
  - American Community Survey (ACS)
  - National Survey of College Graduates (NSCG)
- Improve efficiency
- Improve quality

# Research topics

## ■ AHS

- Tax amount
- Unit size
- Lot size
- Year built
- Year acquired
- Stories
- Bathroom
- Bedrooms
- Total rooms

## ■ NSCG

- Earnings
- Degree
- Field of study
- Employer
- Industry
- Occupation

# Using administrative records in household surveys

- Reduce burden
- Improve efficiency
  - best time of day
  - best mode
  - frame (National Survey of Children's Health, National Teacher and Principal Survey (NTPS))
  - tracing
- Improve quality

# Using administrative records in household surveys

- Reduce burden
- Improve efficiency
- Improve quality
  - AHS, NSCG
  - Survey of Income and Program Participation
  - Consumer Expenditure Survey

# Authority

- Authority to access records

***§6. Information from other Federal departments and agencies; acquisition of reports from other governmental and private sources***

(a) The Secretary, whenever he considers it advisable, may call upon any other department, agency, or establishment of the Federal Government, or of the government of the District of Columbia, for information pertinent to the work provided for in this title.

(b) The Secretary may acquire, by purchase or otherwise, from States, counties, cities, or other units of government, or their instrumentalities, or from private persons and agencies, such copies of records, reports, and other material as may be required for the efficient and economical conduct of the censuses and surveys provided for in this title.

(c) To the maximum extent possible and consistent with the kind, timeliness, quality and scope of the statistics required, the Secretary shall acquire and use information available from any source referred to in subsection (a) or (b) of this section **instead of conducting direct inquiries.**

# Data sources

- Examples of data sources
  - IRS, SSA, HUD, CMS, HHS
  - SNAP, WIC, TANF
  - Property value, taxes, deeds
- Previously collected information
  - Demographic characteristics
  - Housing structure characteristics

# Methods

- Modeling (SIPP)
- Substitution (AHS, NTPS)
- Hybrid approaches



# Research with the ACS

Variable	Question Number
Phone Service	H8g
Year Built	H2
Part of Condominium	H16
Tenure	H17
Property Value	H19
Real Estate Taxes	H20
Have mortgage/mortgage amount	H22a and H22b
Second mortgage/HELOC and payment	H23a and H23b
Sale of Agricultural Products	H5
Social Security	P47d
Supplemental Security Income (SSI)	P47e
Wages	P47a
Self Employment Income	P47b
Interest/dividends	P47c
Pensions	P47g
Residence 1 year ago and Address	P15a
Number of Rooms and Bedrooms	H7a and H7b
Facilities	H8a, H8b, H8c, H8d, H8e, H8f
Fuel type	H13
Acreage	H4

Source: Agility in Action: A Snapshot of Enhancements to the American Community Survey, 2015

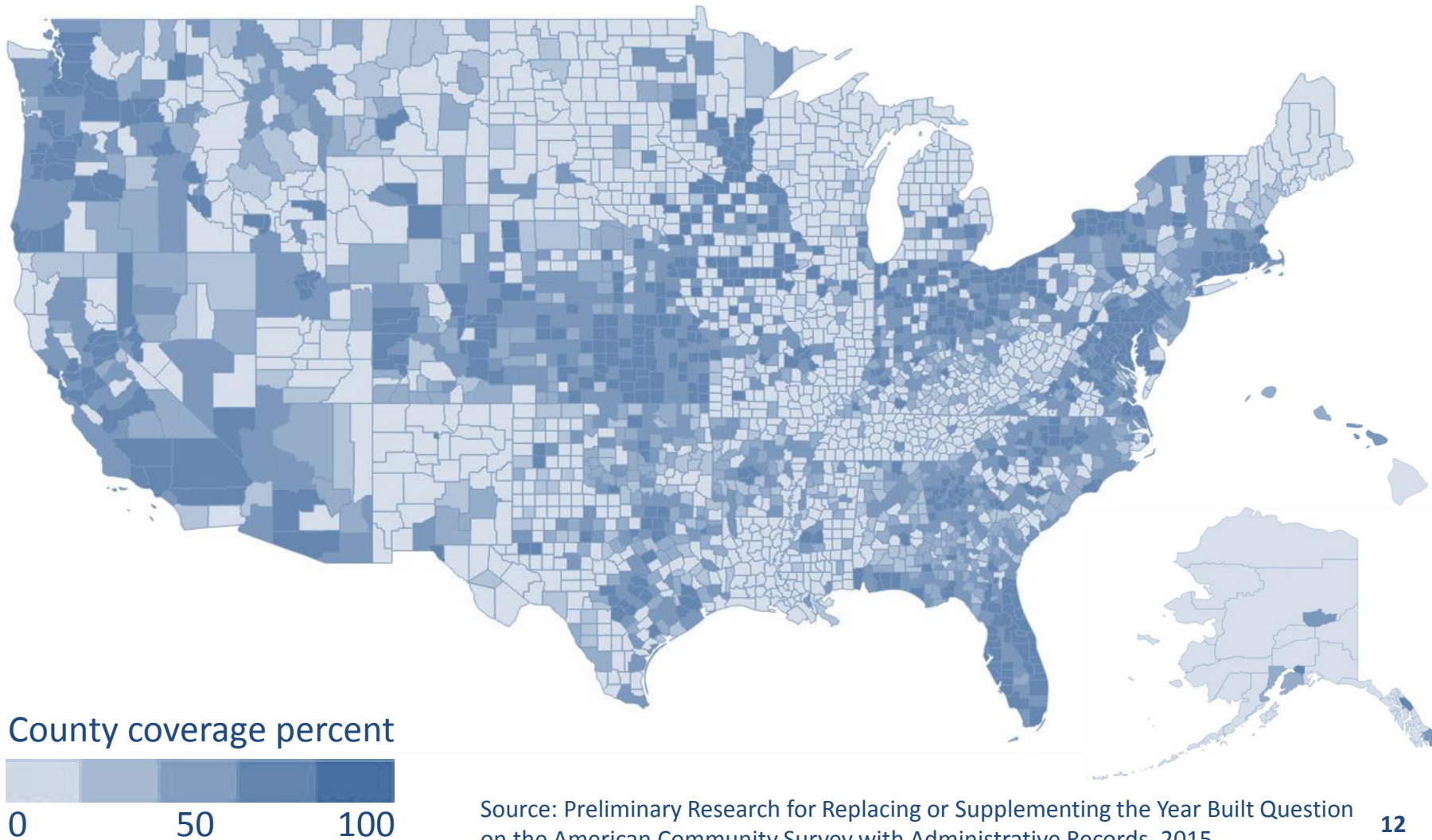
# Burden for ACS questions in study

Variable	Seconds	Variable	Seconds
<b>Wages</b>	41	Real Estate Taxes	9
<b>Interest/dividends</b>	20	Supplemental Security Income	8
<b>Residence 1 year ago and Address</b>	18	Self Employment Income	8
<b>Fuel type</b>	14	Pensions	8
<b>Number of Rooms and Bedrooms</b>	13	Facilities	6
<b>Year Built</b>	11	Second mortgage/HELOC and payment	5
<b>Tenure</b>	11	Acreage	5
<b>Property Value</b>	11	Part of Condominium	4
<b>Have mortgage/mortgage amount</b>	11	Phone Service	1
<b>Social Security</b>	10	Sale of Agricultural Products	1

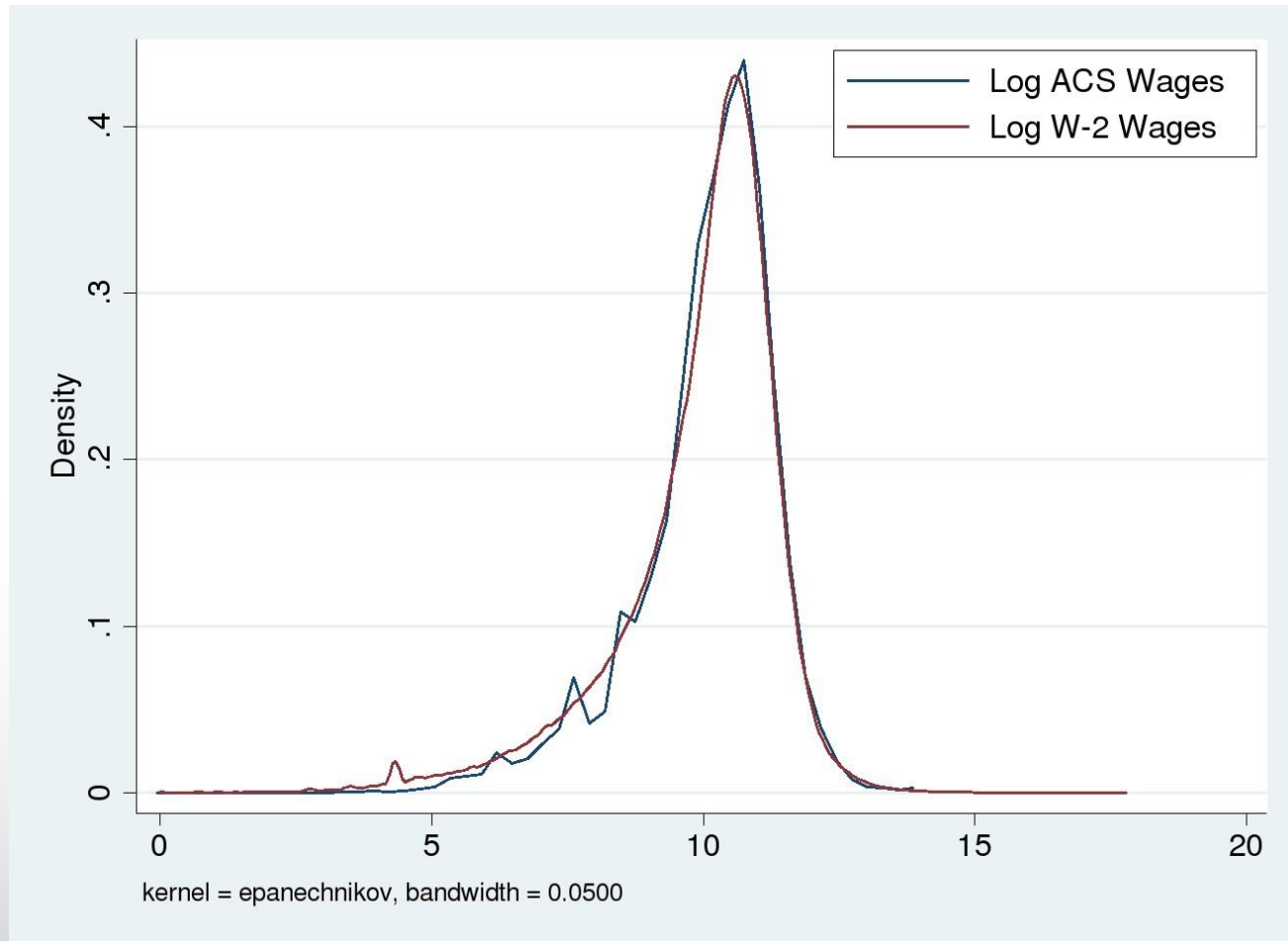
# Difficulty/sensitivity of ACS questions in study

Difficult or Sensitive Variables		Additional Variables in Study
<b>Number of Rooms and Bedrooms</b>	Difficult	Acreage
<b>Property Value</b>	Difficult	Fuel type
<b>Real Estate Taxes</b>	Difficult	Have mortgage/mortgage amount
<b>Year Built</b>	Difficult	Part of Condominium
<b>Facilities</b>	Sensitive	Phone Service
<b>Interest/dividends</b>	Sensitive	Residence 1 year ago and Address
<b>Pensions</b>	Sensitive	Sale of Agricultural Products
<b>Social Security</b>	Sensitive	Second mortgage/HELOC and payment
<b>Supplemental Security Income</b>	Sensitive	Self Employment Income
<b>Wages</b>	Sensitive	Tenure

# Year built coverage by county ACS 2012-Corelogic BSA match



# Income



# Very preliminary FY2016 work

- Housing variables
  - Acreage present for 91%
  - Total Property Value present for 85%
  - Real Estate Tax present for 83%
  - Number of Bedrooms present for 61%
  - Number of Rooms present for 44%

(Source: 2014 ACS housing units matched to TPD)

# Challenges

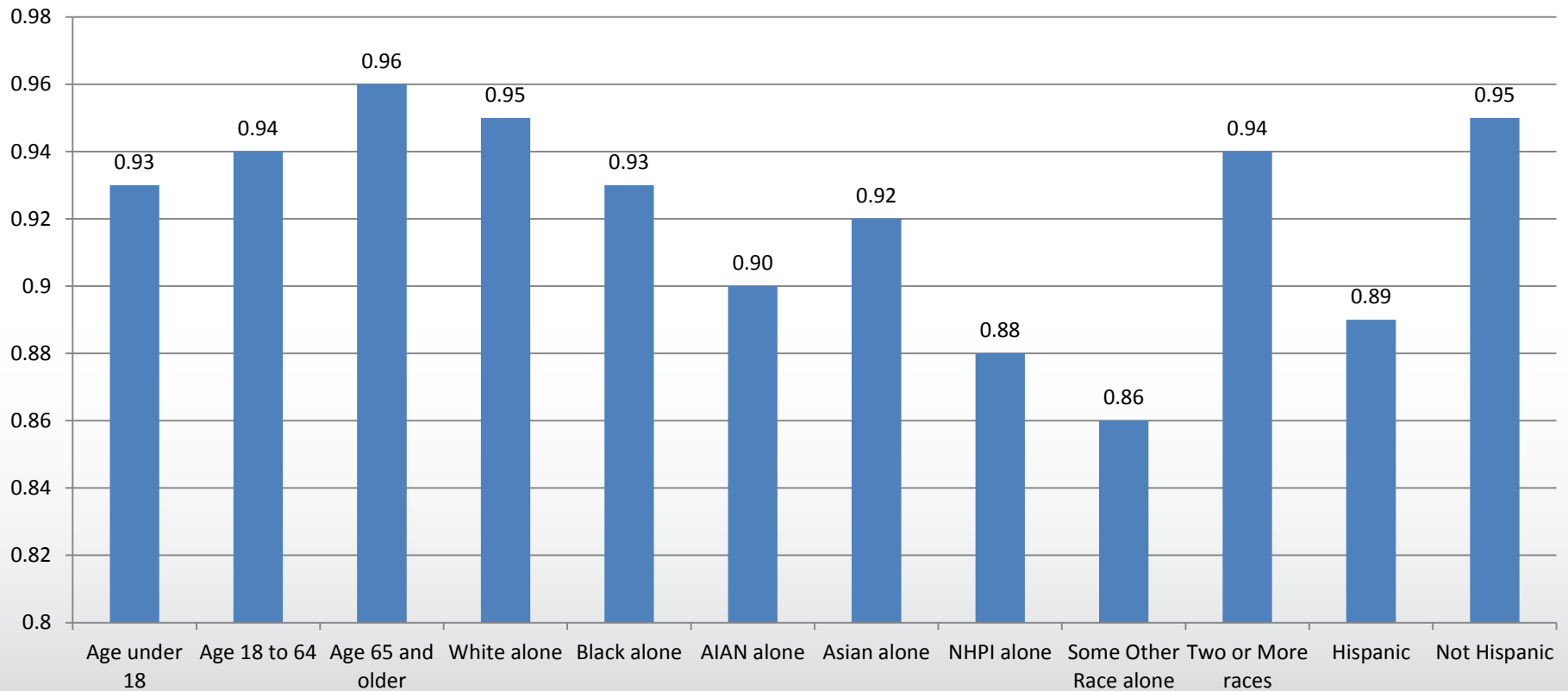
- Data quality
  - How to assess (what is “truth”)
  - Coverage (use matrix sampling where lacking)
  - Concept alignment (last year’s income vs. last 12 month income)
- Matching
- Access and use issues

# Challenges

- Data quality
- Matching
  - Mismatches and missed matches
  - Person and address matching
- Access and use issues

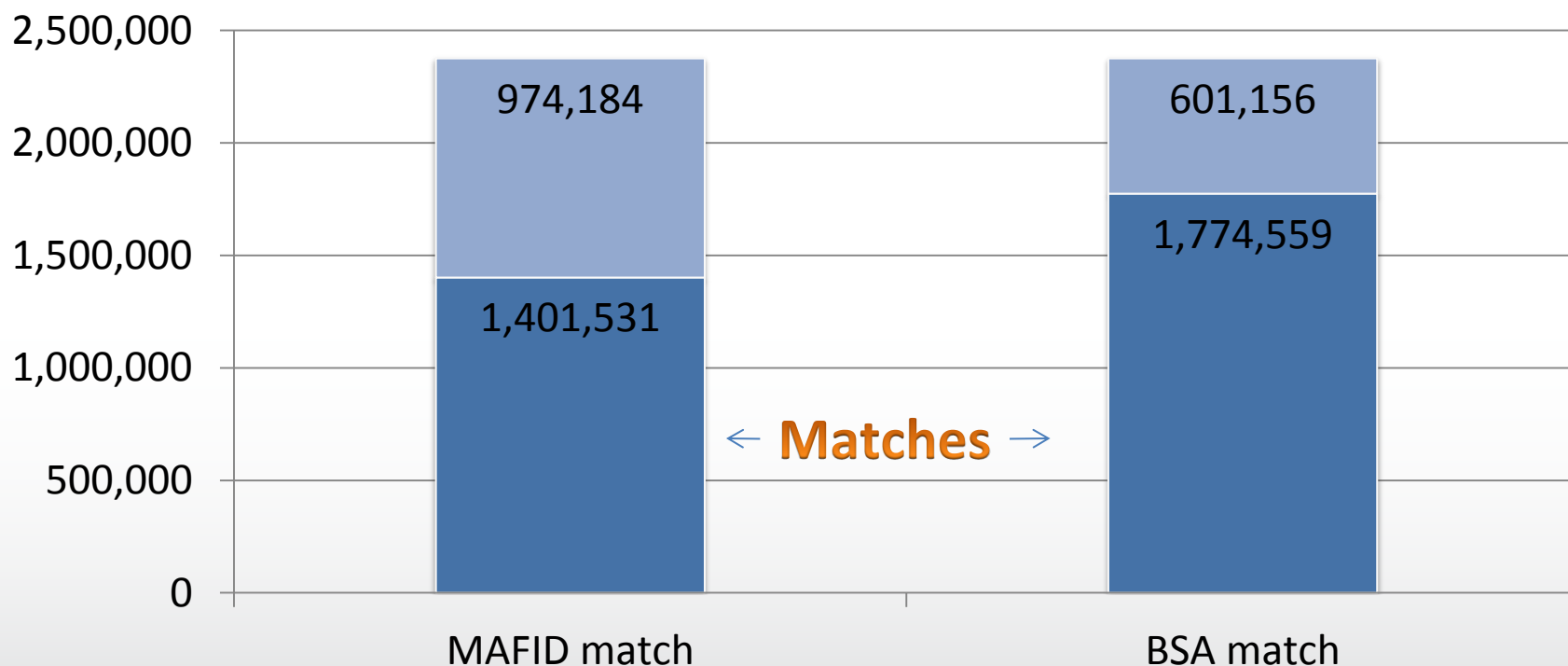


# ACS 2011 PIK rates



Source: Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records, 2016

# ACS 2012-Corelogic MAF and BSA matching

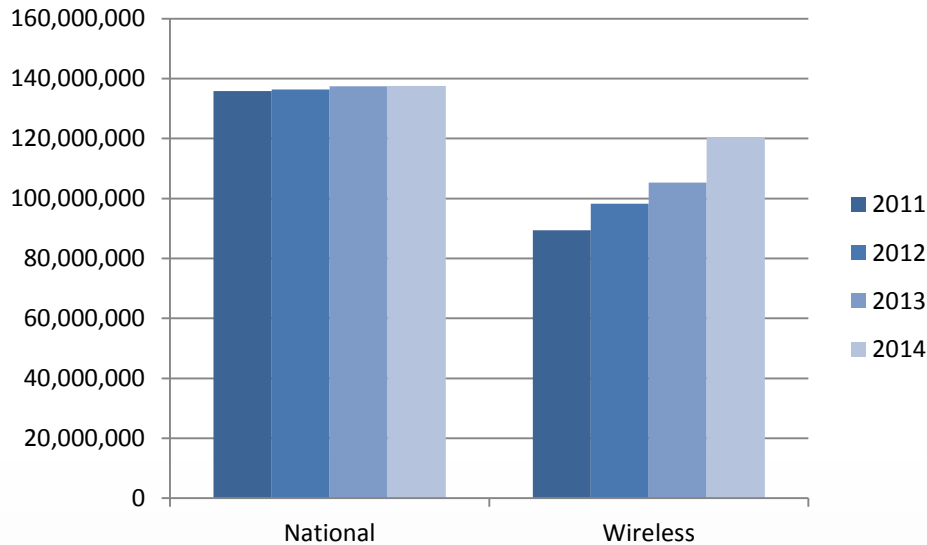


Source: Preliminary Research for Replacing or Supplementing the Year Built Question on the American Community Survey with Administrative Records, 2015

# Challenges

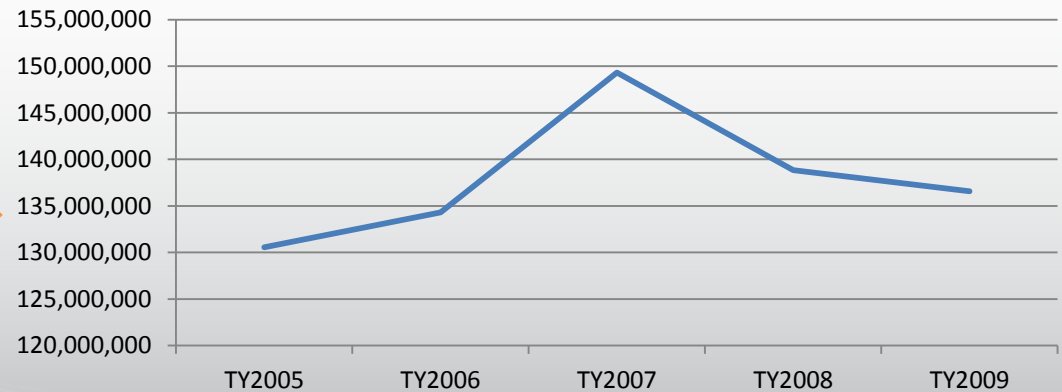
- Data quality
- Matching
- Access and use issues
  - Source volatility and reliability
  - Alignment of reference periods

# Volatility

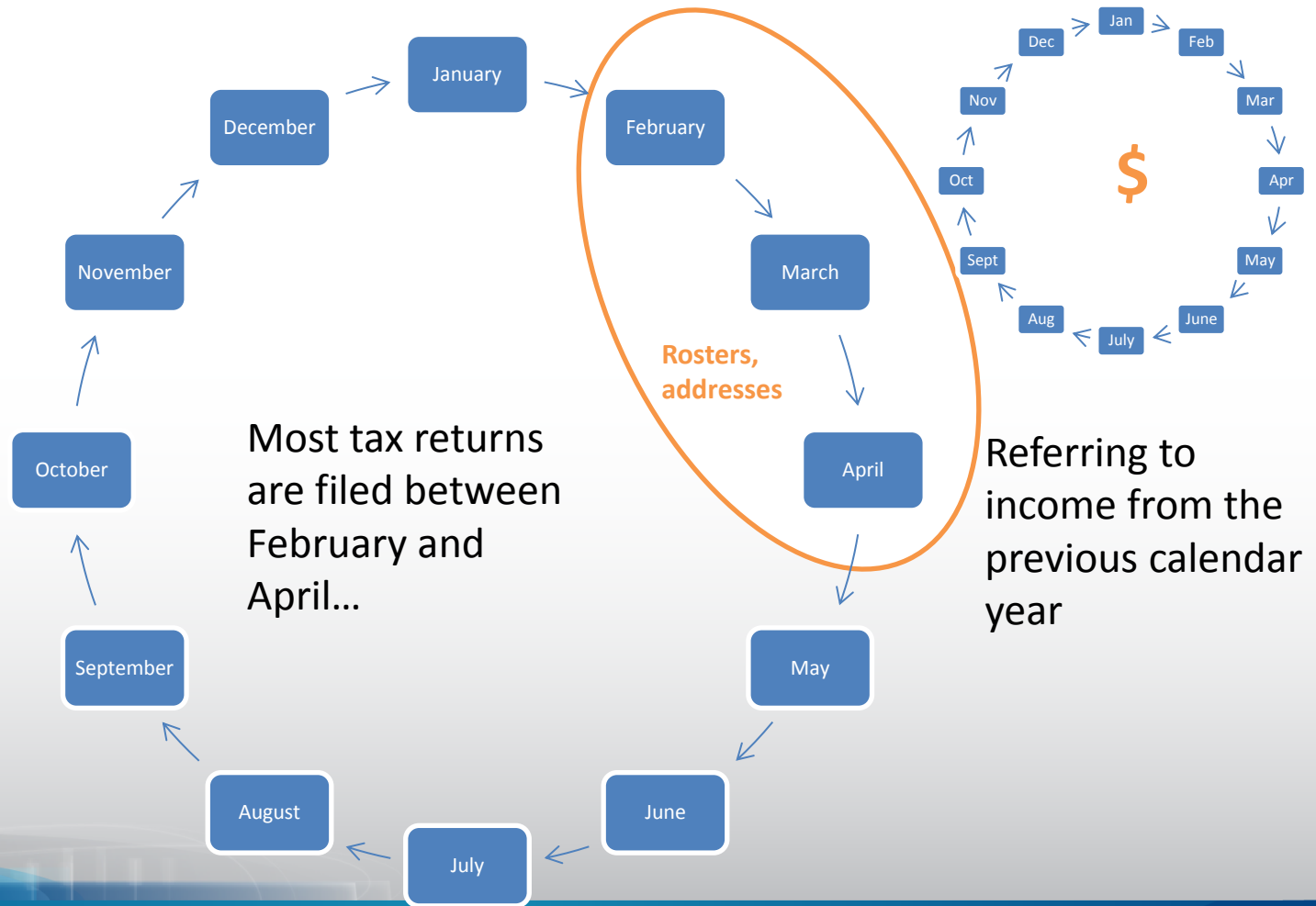


National person file from third party vendor stable, but wireless file is growing over time.

Spike in TY2007 individual income tax returns due to economic stimulus rebate



# Reference periods



# Are we trading burden reduction for lower data quality

- How are we assessing this?
- What else could we do? – study data quality
  - Example - income issues
    - Income not reported to IRS
    - Previous twelve months, annual, or quarterly
    - Do we need eight components or could we use AGI
  - Blended data vs. substitution vs. as-is

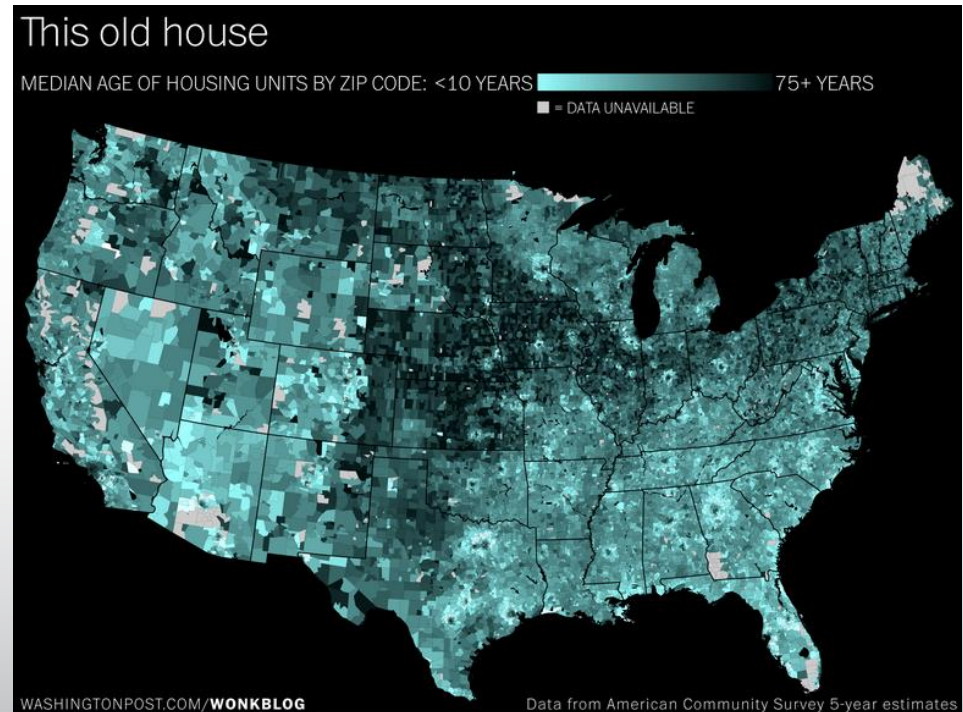
# Thought experiment: Could we implement records in January 2018?

- What needs to happen if we wanted to use **year built** from external source?
  - Prepare data for operational uses
  - Obtain OMB approval for the questionnaire changes
  - Federal register notice before the end of 2016, given our development and print contract schedules
  - Revise ACS edit and imputation processes
  - Simulate impact on 1/5 year products
  - Ensure federal agency buy-in

# Thought experiment: Impact of implementing records

- Impact on releases and products
- Impact on users

- PUMS users →
- HUD reports, buy-in  
(85% match rate for AHS  
single family homes to  
vendor data)





# Questions?