

Big Data Meets Big Survey: Integrating Administrative Records into the American Community Survey

Jennifer M. Ortman and Sandra L. Clark
American Community Survey Office

European Survey Research Association
May 16-19, 2019



U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
[census.gov](https://www.census.gov)

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, operational, or technical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Agenda

- Introduction to the ACS
- Using Administrative Records (AR): Why? When? How?
- Case Study: Replacing Housing Items with AR on the ACS
- Path Forward


The American Community Survey

The American Community Survey is on the leading edge of survey design, continuous improvement, and data quality

- The nation's most current, reliable, and accessible data source for local statistics on critical planning topics such as age, children, veterans, commuting, education, income, and employment
- Surveys **3.5 million** addresses and informs over **\$675 billion** of Federal government spending each year
- Covers **40+ topics**, supports over **300** evidence-based Federal government uses, and produces **11 billion** estimates each year

13199013

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU


 **THE American Community Survey**

This booklet shows the content of the American Community Survey questionnaire.

Start Here

Respond online today at:
<https://respond.census.gov/acs>
OR
Complete this form and mail it back as soon as possible.

This form asks for information about the people who are living or staying at the address on the mailing label and about the house, apartment, or mobile home located at the address on the mailing label.

 **If you need help or have questions** about completing this form, please call 1-800-354-7271. The telephone call is free.

Telephone Device for the Deaf (TDD): Call 1-800-582-8330. The telephone call is free.

¿NECESITA AYUDA? Si usted habla español y necesita ayuda para completar su cuestionario, llame sin cargo alguno al 1-877-833-5625. Usted también puede completar su entrevista por teléfono con un entrevistador que habla español. O puede responder por Internet en: <https://respond.census.gov/acs>

For more information about the American Community Survey, visit our website at: <http://www.census.gov/acs>

➔ Please print today's date.
Month Day Year

➔ Please print the name and telephone number of the person who is filling out this form. We will only contact you if needed for official Census Bureau business.

Last Name

First Name MI

Area Code + Number -


➔ How many people are living or staying at this address?
• **INCLUDE** everyone who is living or staying here for more than 2 months.
• **INCLUDE** yourself if you are living here for more than 2 months.
• **INCLUDE** anyone else staying here who does not have another place to stay, even if they are here for 2 months or less.
• **DO NOT INCLUDE** anyone who is living somewhere else for more than 2 months, such as a college student living away or someone in the Armed Forces on deployment.

Number of people

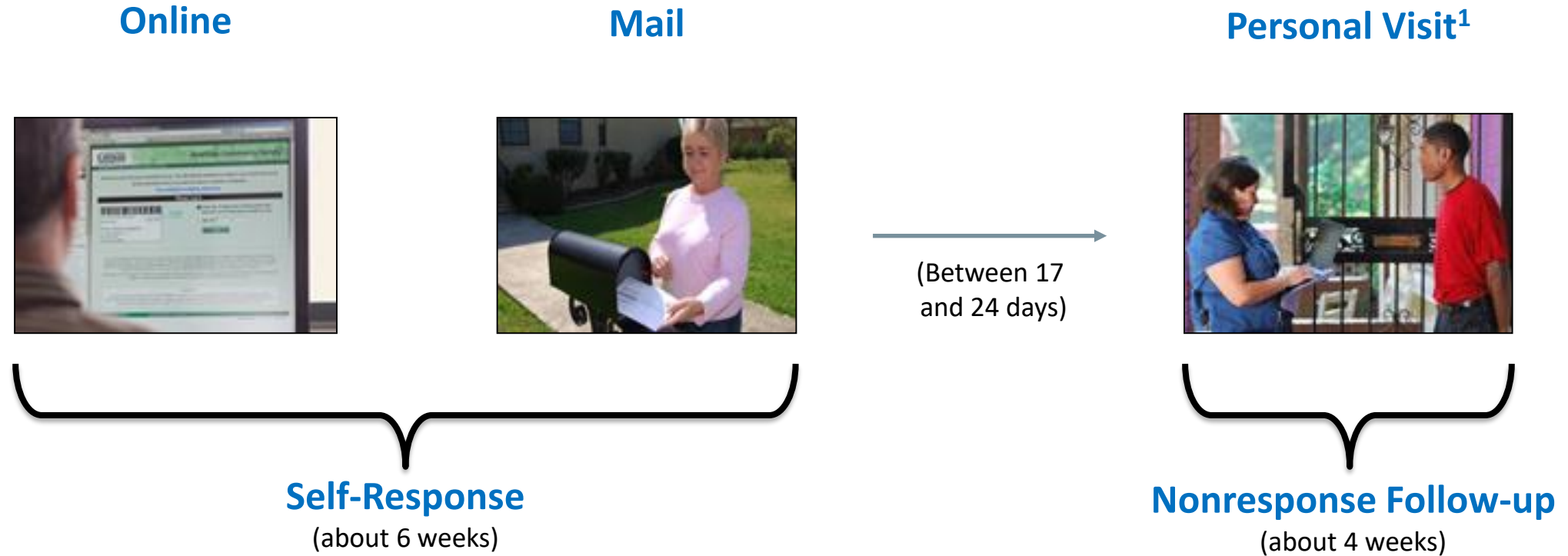
➔ Fill out pages 2, 3, and 4 for everyone, including yourself, who is living or staying at this address for more than 2 months. Then complete the rest of the form.

FORM ACS-1(INFO)(2019)
OMB No. 0607-0810
OMB No. 0607-0936

INFORMATION COPY



ACS Data Collection Process



¹ The Personal Visit is conducted via Computer Assisted Personal Interviewing (CAPI).

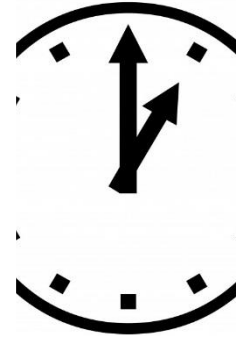
ACS Program Priorities



Why and How to Use Administrative Data?



Increase data quality by filling in missing responses and using AR to evaluate and enrich survey data



Save time and improve respondent experience by reducing the number of questions asked on the ACS



Provide cost savings by identifying vacant housing units and reducing the need for followup visits



Mandated by Title 13 of the U.S. Code to use already available information

Charting the Course

Future Research

Expanding Horizons:
Improving Data
Quality and
Products

Identifying AR
Sources

Previous Research

Evaluating AR
Sources With
Guiding
Principles

Cycle continues as we
obtain new AR sources

Current Research

Leveraging AR to
Model Survey
Item Responses

Incorporate in
Edit and
Imputation
Procedures

Direct
Replacement
Simulation

Administrative Sources

Federal data

- U.S. Census Bureau
- Internal Revenue Service
- Housing and Urban Development
- Childcare Development Fund
- Medicaid and Medicare
- Social Security Administration
- Veteran's Affairs
- U.S. Postal Service
- Selective Service

State and Local data

- Women, Infants, and Children
- Temporary Assistance for Needy Families
- Supplemental Nutrition Assistance Program
- Child Care Subsidy
- Public school districts

Third Party data

- Corelogic property and tax foreclosure
- VSGI consumer households

12 Guiding Principles

Authorization

Availability

Conceptual Alignment

Coverage

Data Source

Disclosure Avoidance

Impacts on Estimates

Intended Use

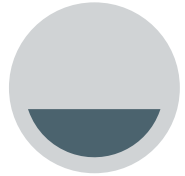
Population Universe

Quality

Reliability

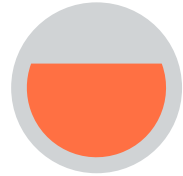
Temporal Alignment

Evaluating Administrative Sources



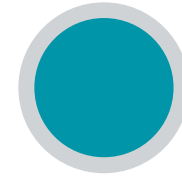
Least Promising

Phone service
Part of condominium
Tenure
Secondary mortgage/amount
Number of rooms/bedrooms
Facilities
Fuel type



Somewhat Promising

Have mortgage
Agricultural sales

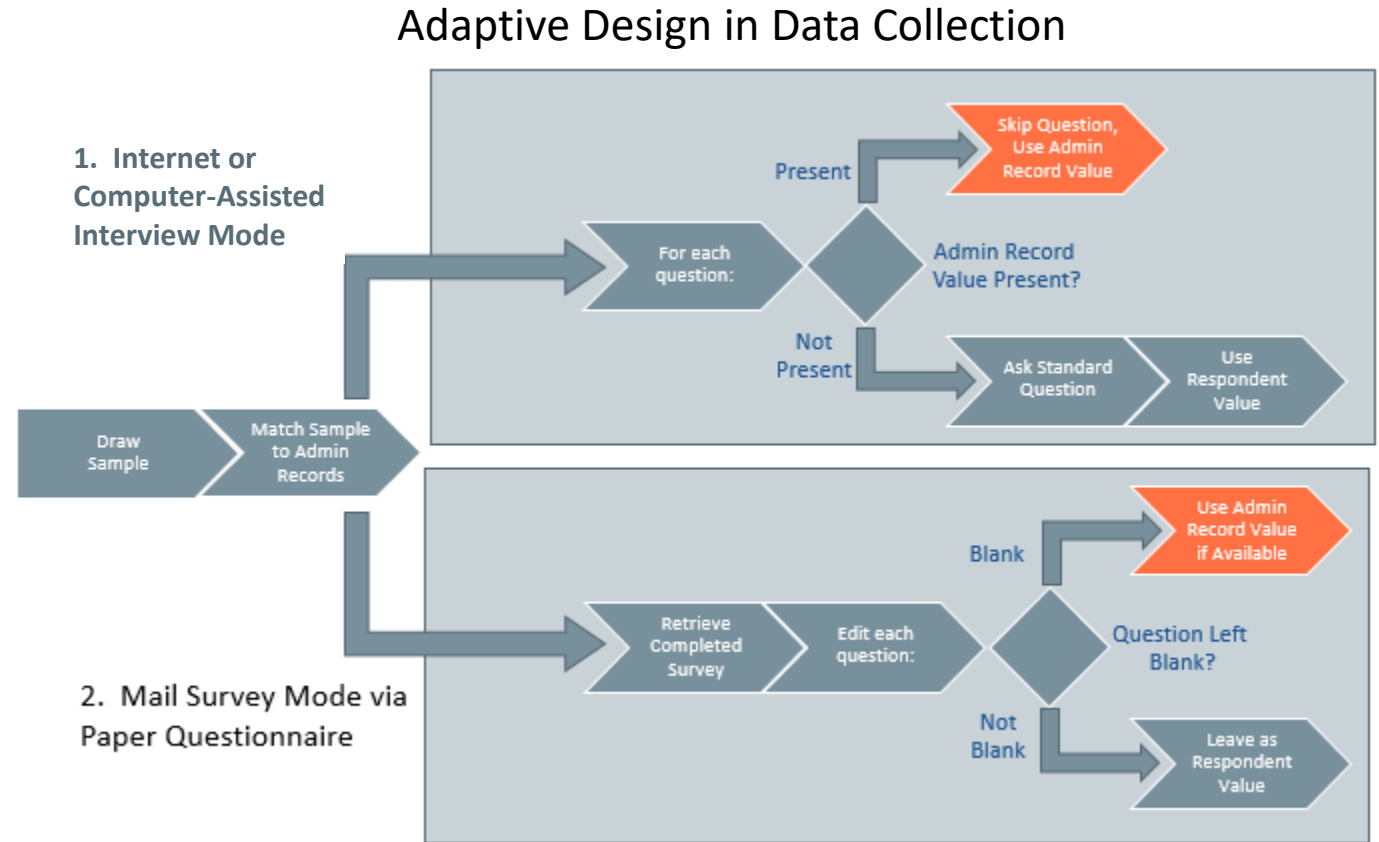


Most Promising

Acreage
Property Value
Real Estate Tax
Year Built

Case Study: Replacing Housing Items with Administrative Records (AR)

- Used 2015 ACS Responses
- Direct substitution for:
 1. Year built
 2. Acreage
 3. Real estate taxes
 4. Property value
- Produced “Simulated” version to compare to “Published” 2015 ACS estimates



Respondents in the internet or computer-assisted modes would not be asked the question; those in the mail mode would still be asked the question on the paper questionnaire

Case Study: Simulated vs Published Estimates

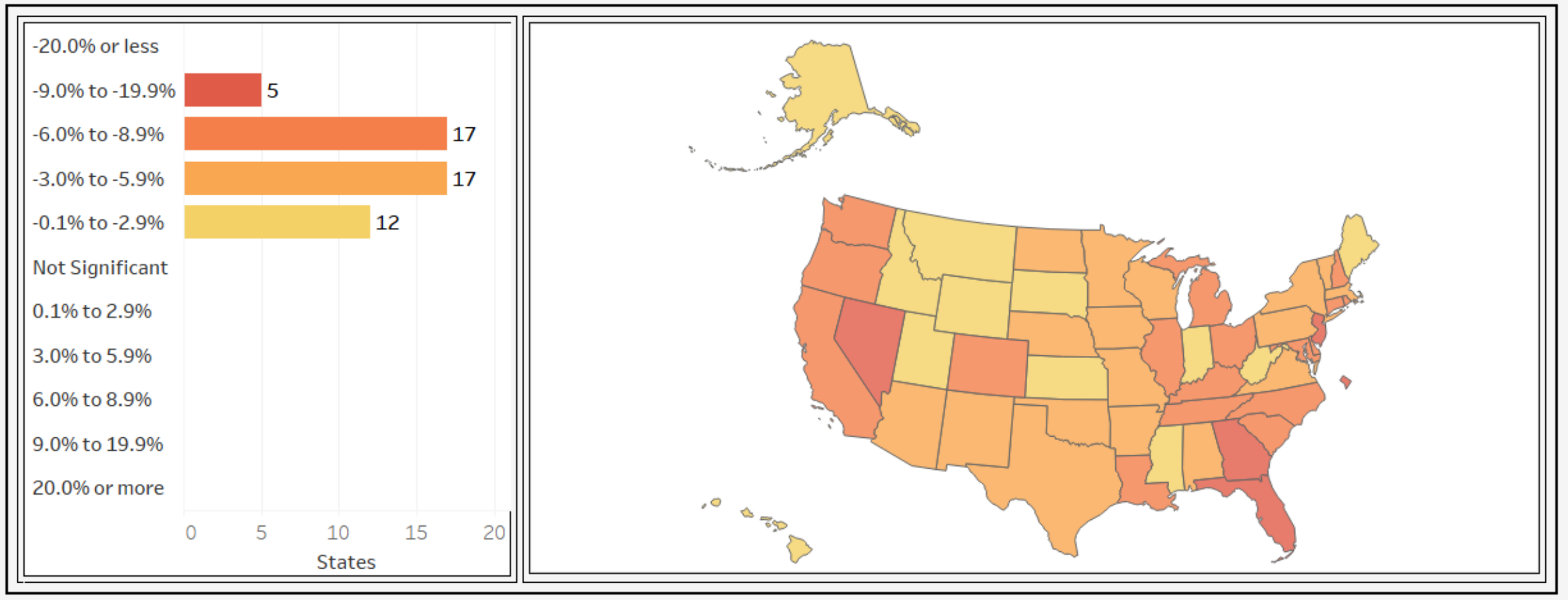
- Majority of the summary metrics and key measures studied were statistically different
- Direction of differences varied, but for many items **Simulated** was lower than **Published**
- **Simulated** item allocation¹ rates significantly lower than **Published**

Item	Simulated	Published	Difference	MOE
Acreage	1.4	3.7	-2.3	0.1
Year Built	12.7	17.8	-5.2	0.2
Property Value	5.0	12.0	-7.0	0.1
Property Tax	4.5	16.9	-12.4	0.1

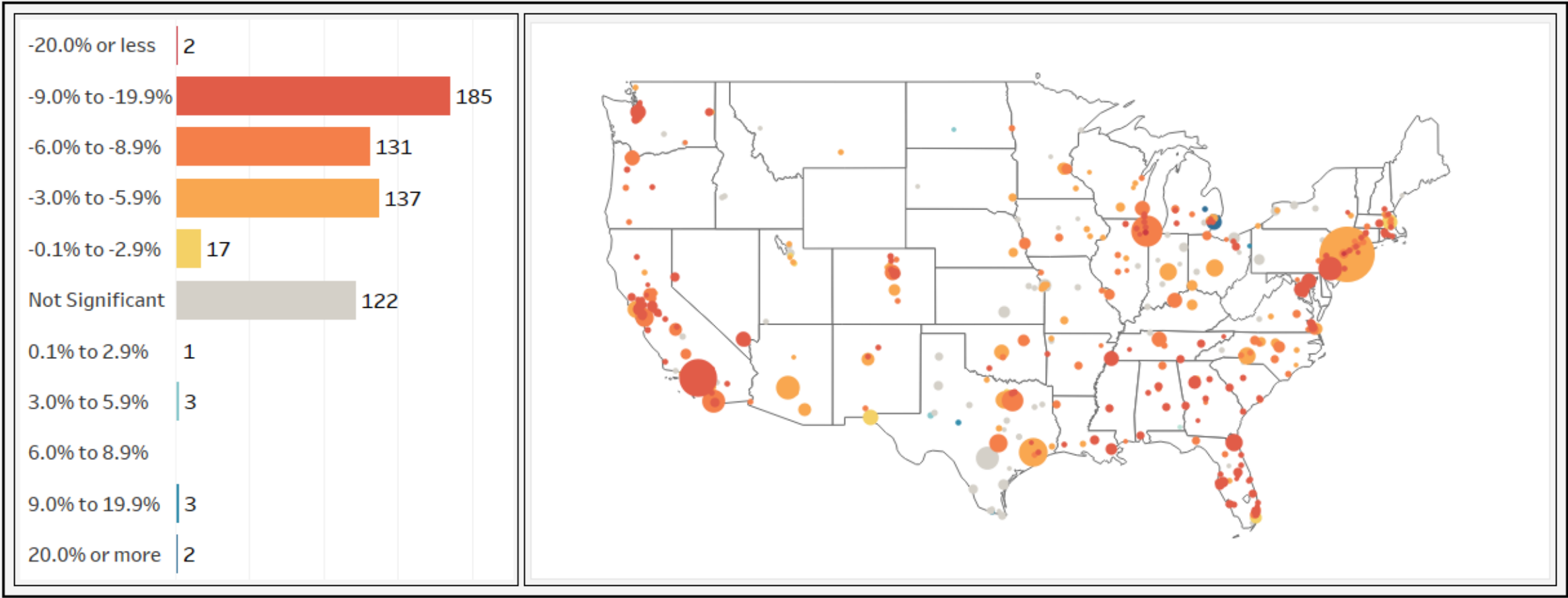
- Impacts other survey items besides 4 test items

¹ An allocation involves using statistical procedures to impute for missing values 12

Case Study: Percent Difference in Median Property Value: Simulated minus Published - State

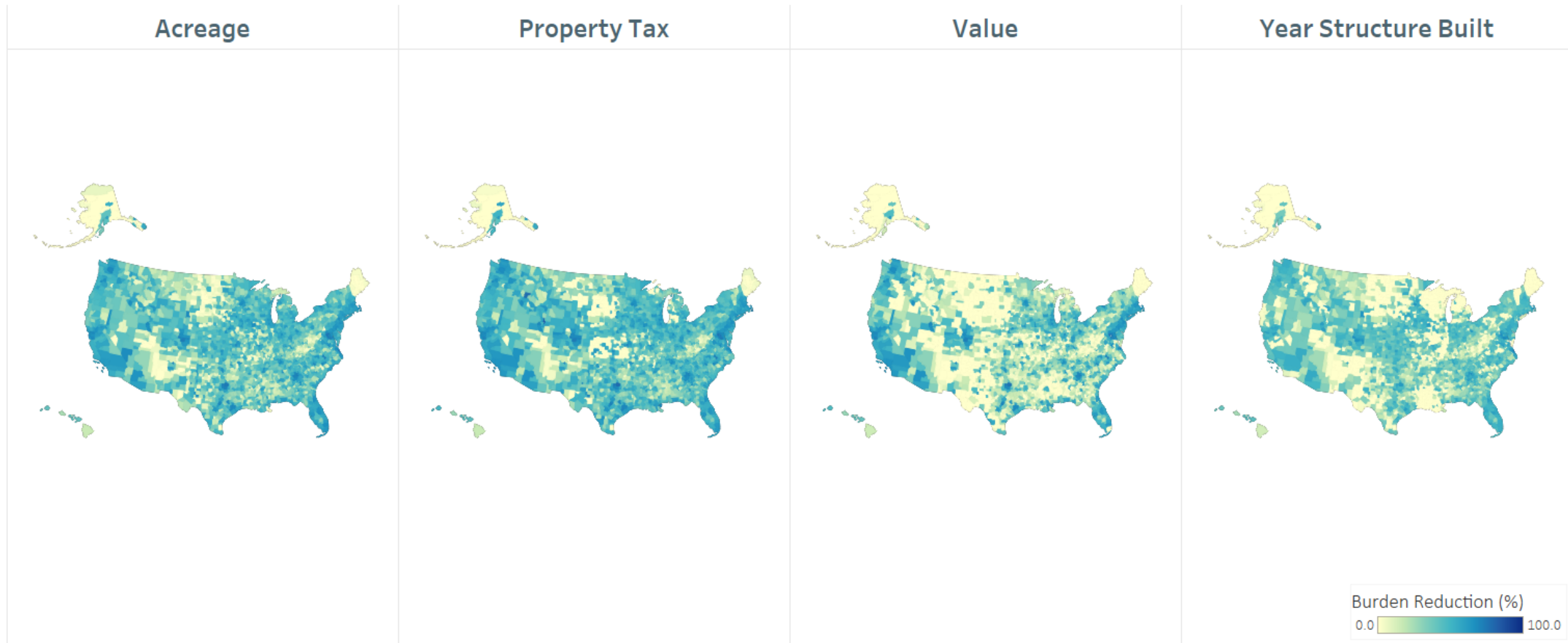


Case Study: Percent Difference in Median Property Value: Simulated minus Published - Place



Note: Estimates were not calculated for counties not shown because they do not meet the 1-year ACS population threshold of 65K or more. Also, Alaska and Hawaii are not shown, but Honolulu, HI and Anchorage, AK both had differences that were not significant.

Case Study: Burden Reduction by County and Survey Question



This visualization shows the percentage of households that would have a reduction in burden because they would not be asked the question using the simple replacement method.

Case Study: Conclusions

Opportunities

- Reduces respondent burden in terms of asking these survey items
- Improves item allocation rates
- Potential for improving edit and imputation methods

Challenges

- AR are different from ACS response data
 - Collection and reporting differs by jurisdiction
 - Time lag between survey year and AR
 - Differences in coverage/availability of AR
 - Break in series
- Implementation challenges
 - Discovered impact on other survey items
 - Difficult to adapt to mail mode

Current Research: Integrating AR in ACS Edit and Imputation Procedures

Use AR data to fill-in missing ACS values for

- Age, Sex, Race, Hispanic Origin

Adapt method developed for 2020 Census using following data sources

- 2010 Census
- 2010-2016 ACS
- Social Security Administration Numident
- Census Best Race/Ethnicity File

Research questions

- How many and what proportion of missing ACS values can be filled in with AR?
- How do simulated and published estimates compare?
- Are there impacts to other survey items?

Current Research: Leveraging AR to Model Survey Data

What was this person's total income during the PAST 12 MONTHS? Add entries in questions 43a to 43h; subtract any losses. If net income was a loss, enter the amount and mark (X) the "Loss" box next to the dollar amount.

None OR Loss

\$, , .00

TOTAL AMOUNT for past 12 months

1040 Department of the Treasury—Internal Revenue Service (99) 2016 CMB No. 1545-0074 IRS Use Only—Do not write or staple in this area

U.S. Individual Income Tax Return 2016

Your first name and initial: Samuel P
Last name: Taxpayer
Your social security number: 123-45-6789

If a joint return, spouse's first name and initial: Felicity Q
Last name: Taxpayer
Spouse's social security number: 432-19-8765

Home address (number and street), if you have a P.O. box, see instructions: 789 Tuxedo Drive
City, town or post office, state, and ZIP code, if you have a foreign address, also complete spaces below (see instructions): Bronxville NY 10708

Filing Status: 1 Single, 2 Married filing jointly (even if only one had income), 3 Married filing separately. Enter spouse's SSN above and full name here, 4 Head of household (with qualifying person), 5 Qualifying widow(er) with dependent child

Exemptions: 6a Yourself, 6b Spouse

Research utility of IRS data for supplementation or replacement
Examine alignment of ACS responses with corresponding IRS values

Develop and test statistical model-based approaches to

- Examine levels of misalignment between ACS and IRS values
- Generate refined estimates for item supplementation
- Simulate various item replacement scenarios

Research questions

- How do ACS responses to questions compare to equivalent information from the relevant IRS sources?
- How are ACS estimates impacted by IRS data item supplementation?
- It is possible to use statistical models to produce more accurate and precise ACS level estimates?

The Path Forward

- Continue research initiatives to:
 - Improve linkage methods
 - Broaden access to administrative data sources
 - Assess the quality of administrative data sources
 - Develop methods to harmonize survey/administrative data and address coverage issues
 - Create blended data products
- Implement in editing and imputing procedures
- Collaborate with others on AR integration and utilization

Thank you!

Jennifer.M.Ortman@census.gov