# Making Administrative Records Key to Operational Agility for the American Community Survey

Jennifer M. Ortman, Victoria A. Velkoff, and Sandra L. Clark
*American Community Survey Office*

Big Data Meets Survey Science (BigSurv18) Conference, Barcelona, Spain

October 25-27, 2018

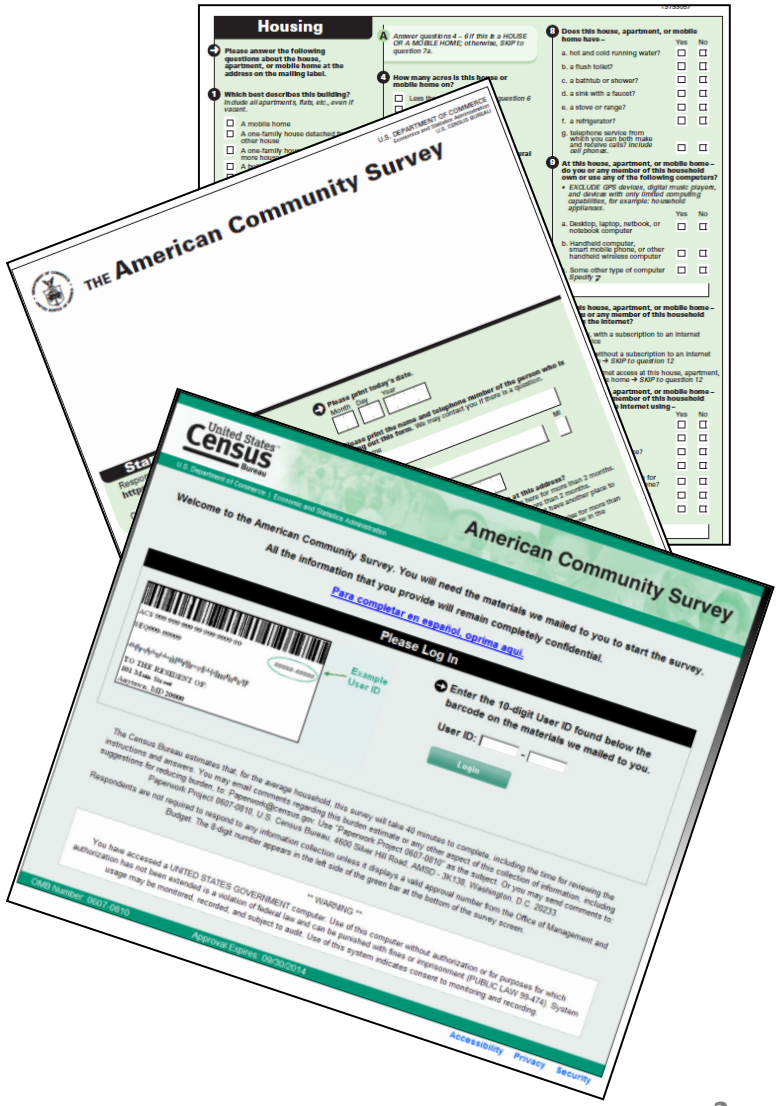
United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

# The American Community Survey

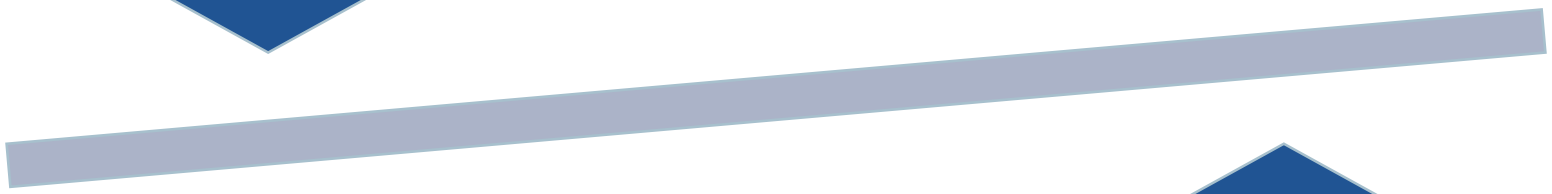
## The Basics

- Ongoing **monthly survey sent to 3.5 million addresses per year** to produce detailed population and housing estimates each year
    - Visit 20,000 Group Quarter facilities and sample approximately 194,000 residents each year
- Designed to **produce critical information on small areas and small population groups** previously collected on the decennial long form
- Covers **35+ topics** and supports over 300 known Federal government uses
- Used to **distribute more than $675 billion federal funds** each year
- Data released annually
    - 1-year estimates (12 months of data)
    - 5-year estimates (60 months of data)



# Striking a Balance

**Declining response rates** as well as growing **concerns about privacy and confidentiality** of data challenge our ability to collect information using surveys.



Society demands **more data** at a **rapid pace** to meet the needs of the changing landscape of America's communities.

# The Promise of Administrative Records

- The Census Bureau is mandated, to the extent possible, to use administrative records (AR) for the efficient and economical conduct of censuses and surveys (Title 13 U.S.C § 6)
- Reduce the amount of information we request from respondents
- Increase data reliability
- Provide cost savings by reducing the need for follow up visits
- Provide additional information to enrich census and survey sources
- Benchmark for evaluating census and survey data
- Support survey operations in remote areas

# Administrative Records at the Census Bureau

## 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

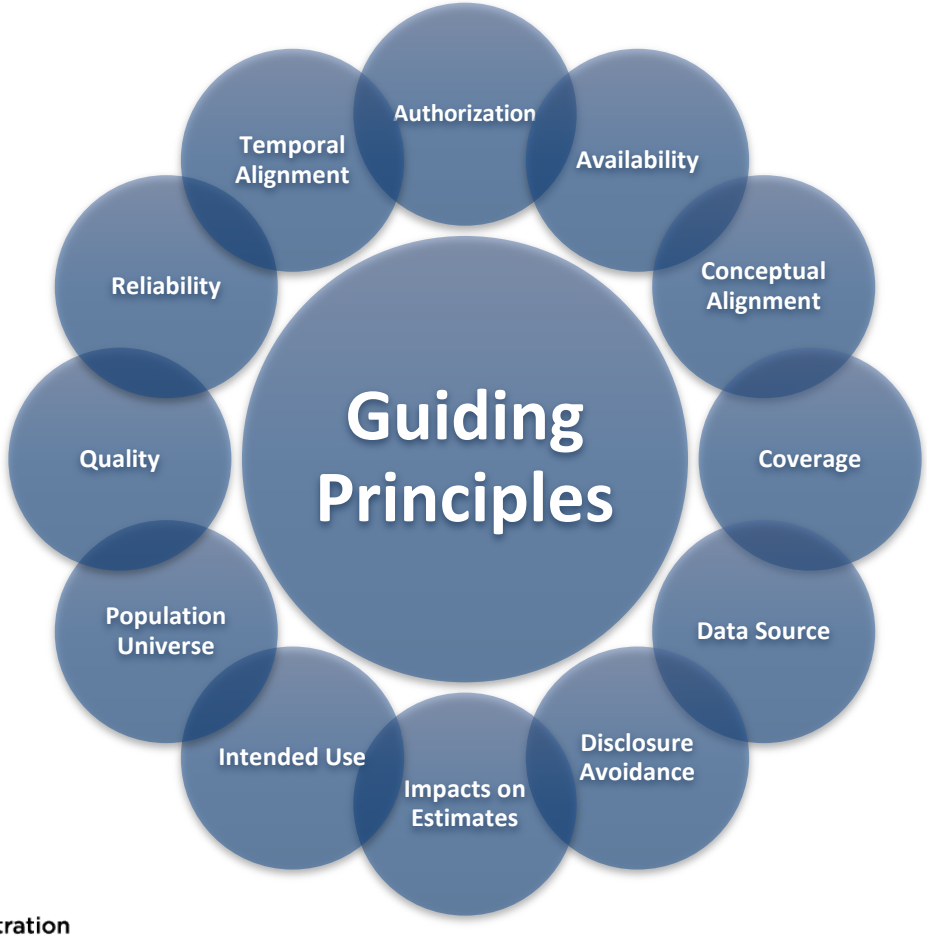
# What Have We Done?

- Established **guiding principles** to determine what administrative data are appropriate for the ACS
- Evaluated the coverage and quality of administrative records (government and third-party sources) to **identify the most promising sources**.
- Tested **direct replacement of ACS housing items** (year built, acreage, property value, and real estate taxes) to evaluate impact on data products.
- Tested **modeling and direct replacement of ACS income** items to evaluate feasibility of replacing or augmenting the ACS income questions.
- Tested the **use of administrative records for item imputation** of race, Hispanic origin, and age for the 2020 Census.

# What Have We Done?

- Established **guiding principles** to determine what administrative data are appropriate for the ACS
- Evaluated the coverage and quality of administrative records (government and third-party sources) to **identify the most promising sources**.
- Tested **direct replacement of ACS housing items** (year built, acreage, property value, and real estate taxes) to evaluate impact on data products.
- Tested **modeling and direct replacement of ACS income** items to evaluate feasibility of replacing or augmenting the ACS income questions.
- Tested the **use of administrative records for item imputation** of race, Hispanic origin, and age for the 2020 Census.

# 12 Guiding Principles



# Most and Least Promising Housing Topics

**Most Promising**

- ✓ Property Value
- ✓ Real Estate Tax
- ✓ Year Built
- ✓ Acreage

**Least Promising**

- o Part of Condominium
- o Phone Service
- o Facilities (Kitchen/Bathroom)
- o Fuel Type
- o Tenure
- o Have a Mortgage
- o First/Secondary Mortgage Payment
- o Number of (Bed)rooms
- o Agricultural Sales

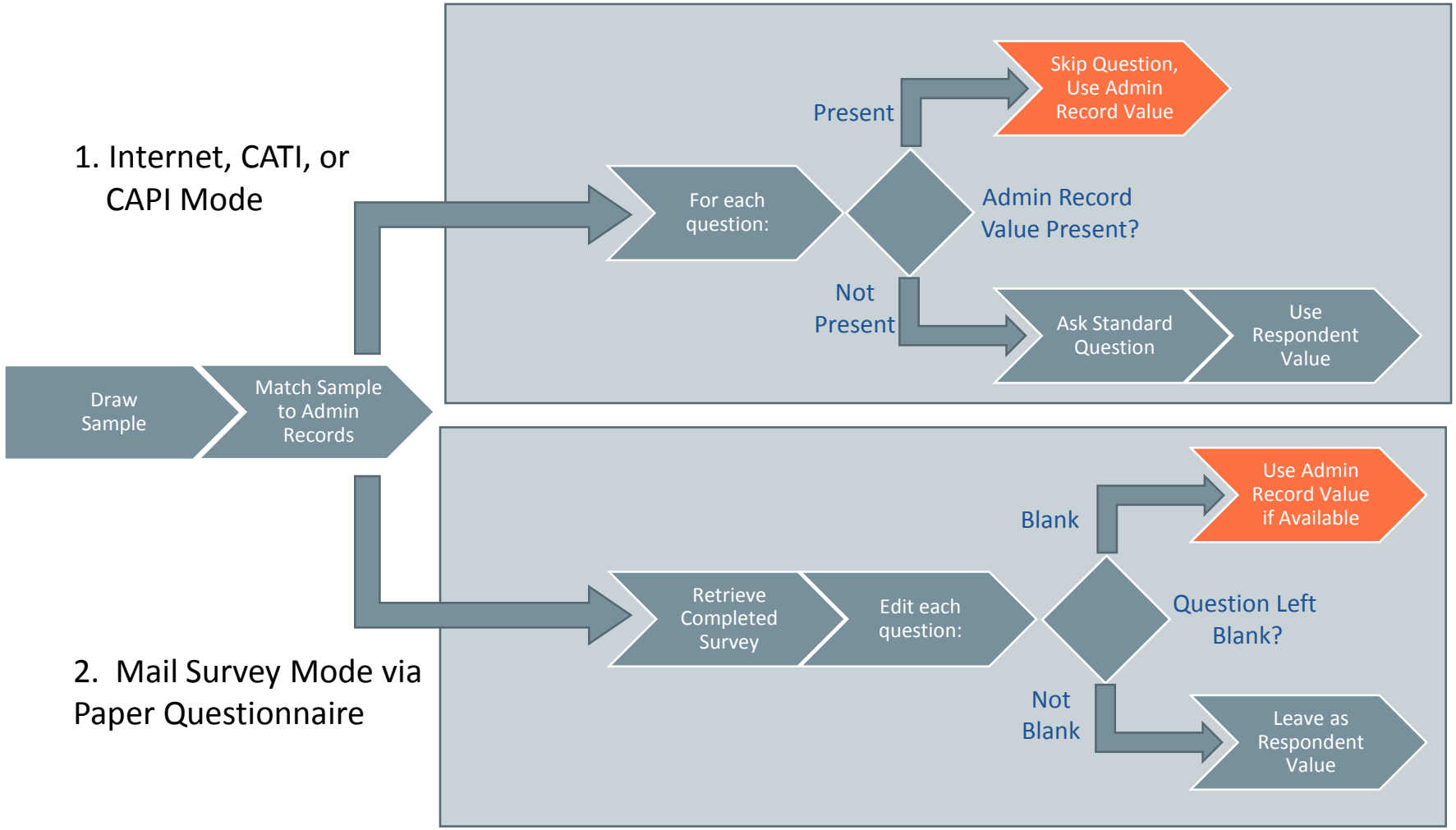
# Case Study:
# Housing Administrative Record Simulation

Study Authors: Sandra L. Clark, R. Chase Sawyer, Amanda Klimek, Christopher Mazur, William Chapin, Ellen Wilson

# Study Objectives

- Replace responses in the 2015 ACS to questions about **year built**, **acreage**, **real estate taxes**, and **property value** with administrative records to:
  - Study impact on estimates and data products
  - Test feasibility of implementing methods
  - Learn effects on production process
  - Satisfy ACS program goals to be reputable, researched, and responsive

# Adaptive Design



# Results

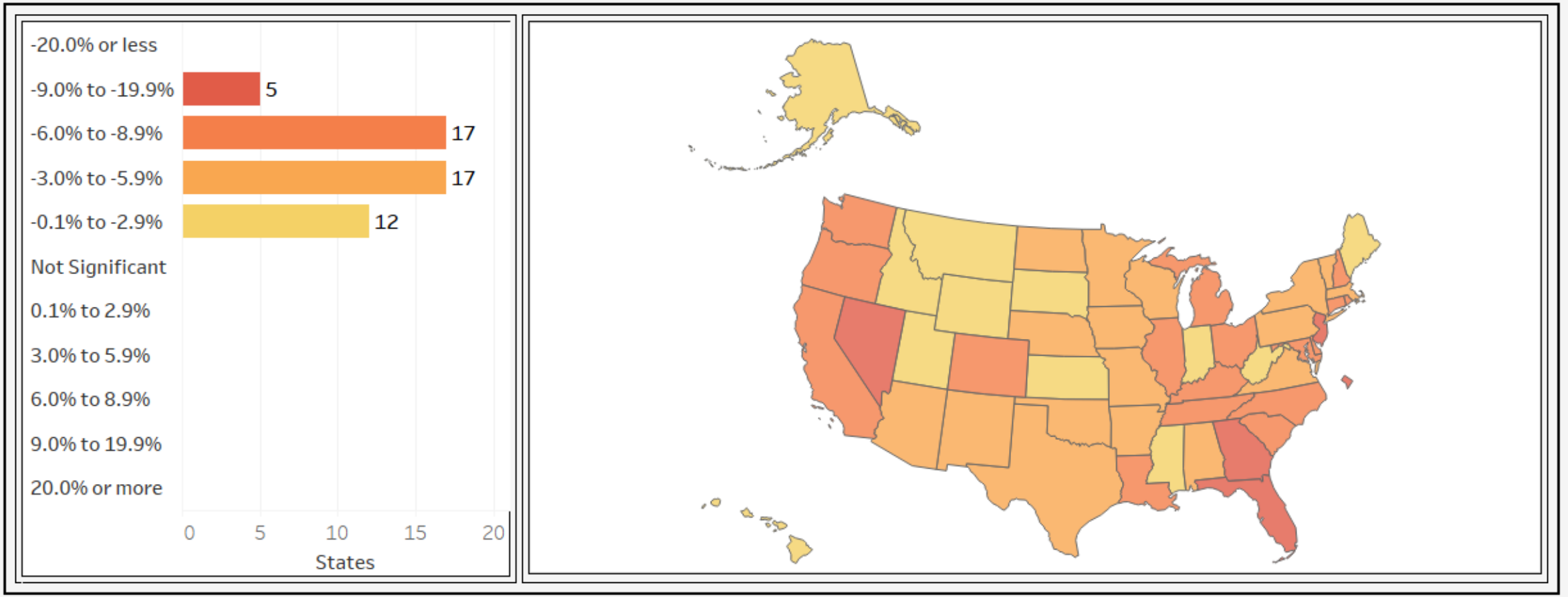
**Summary metrics and key measures were compared.**

- Approx. **79%** of 575 U.S.-level estimates were statistically different
- Direction of differences varied, but for many items **Simulated** was lower than **Published**
- **Simulated** item allocation rates significantly lower than **Published**

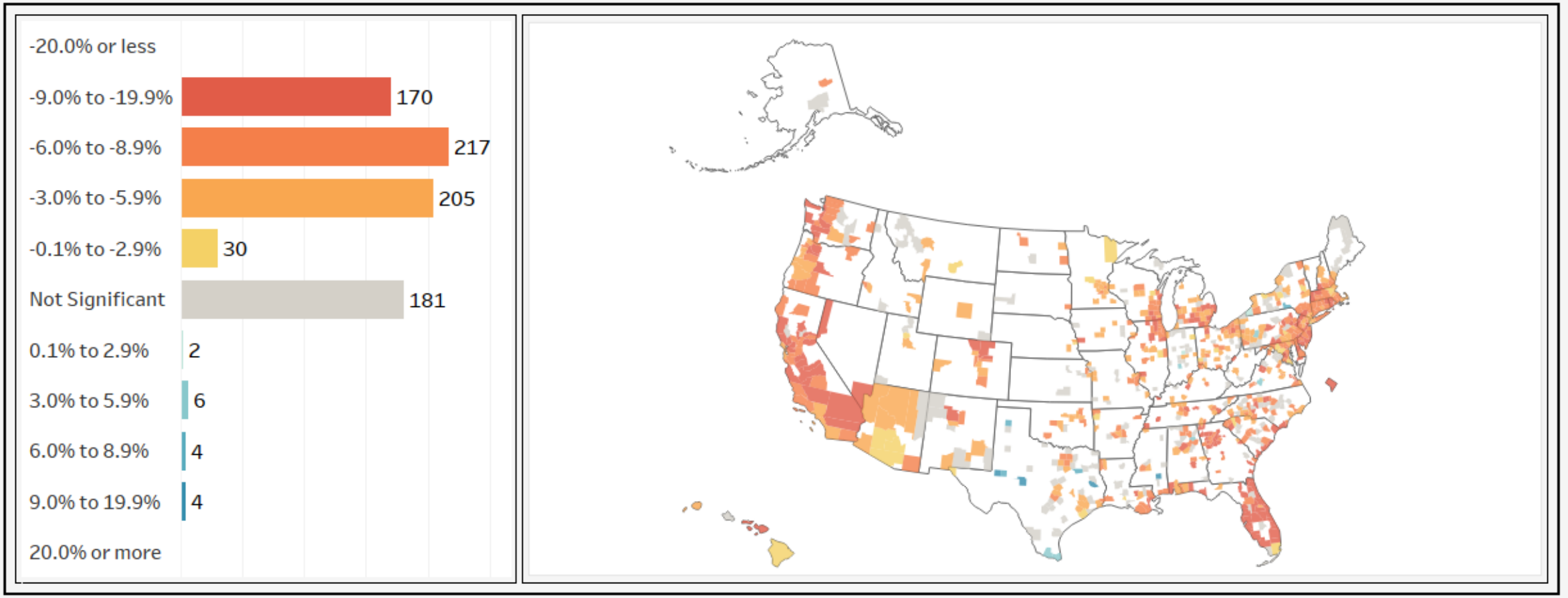
| Item | Simulated | Published | Difference | MOE |
|---|---|---|---|---|
| Acreage | 1.4 | 3.7 | **-2.3** | 0.1 |
| Year Built | 12.7 | 17.8 | **-5.2** | 0.2 |
| Property Value | 5.0 | 12.0 | **-7.0** | 0.1 |
| Property Tax | 4.5 | 16.9 | **-12.4** | 0.1 |

- Impacts other survey items besides 4 test items

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

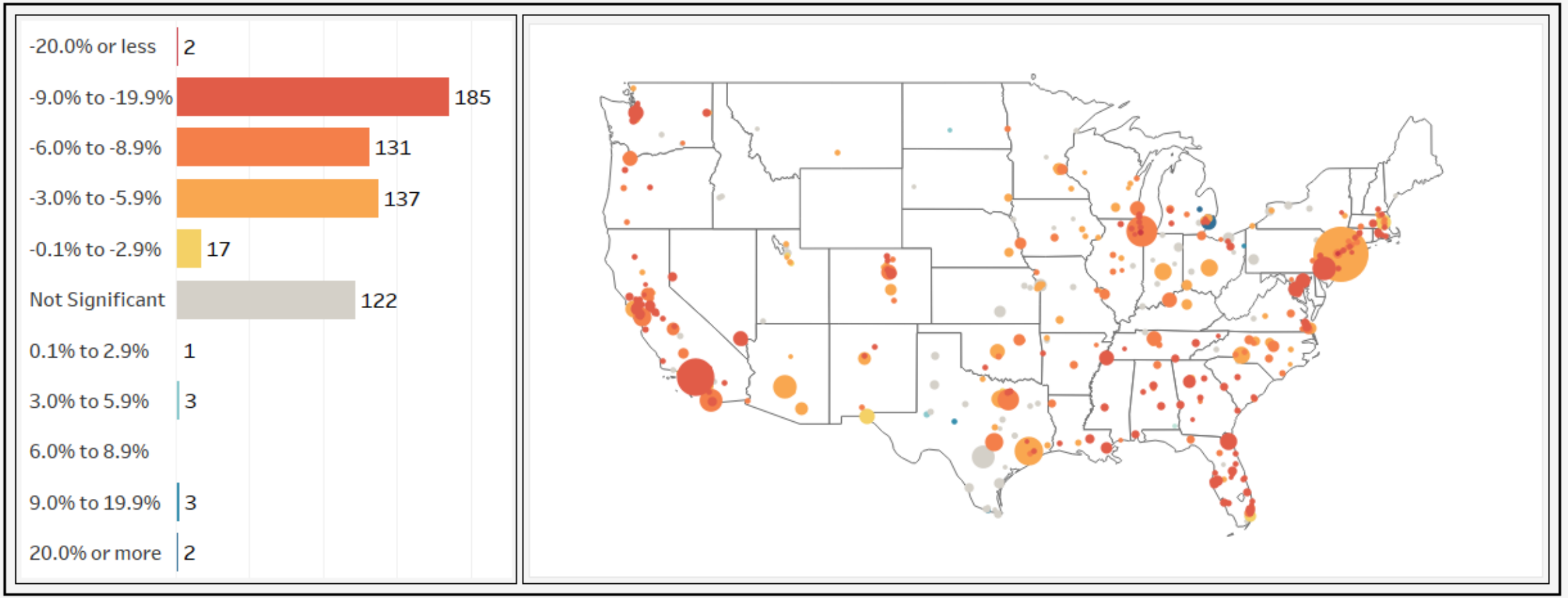


# Percent Difference in Median Property Value: Simulated minus Published - County




Note: Estimates were not calculated for counties not shown because they do not meet the 1-year ACS population threshold of 65K or more.

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



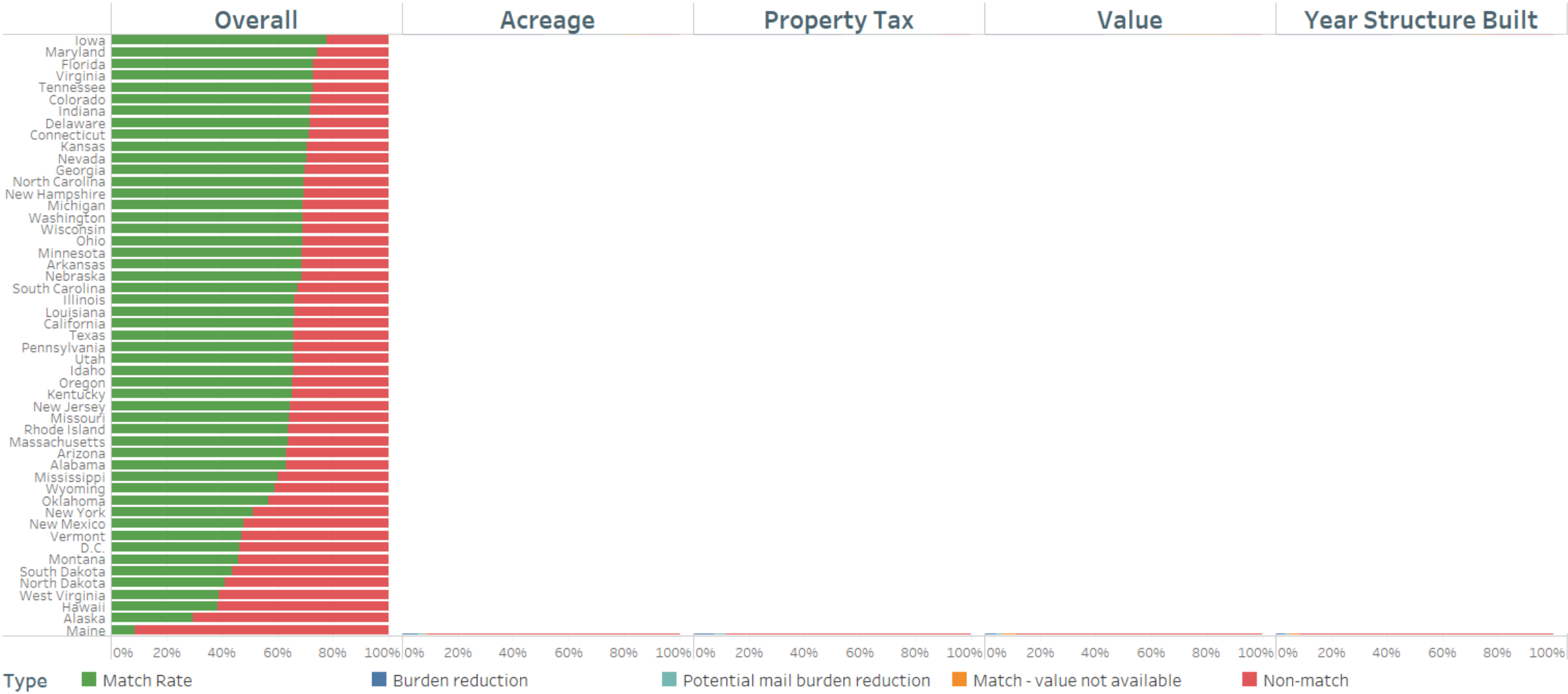

Note: Estimates were not calculated for counties not shown because they do not meet the 1-year ACS population threshold of 65K or more. Also, Alaska and Hawaii are not shown, but Honolulu, HI and Anchorage, AK both had differences that were not significant.

# Match Status

**Match** = ACS responding household with a AR value for at least one of the 4 tested items

**Non-match** = ACS responding household that did NOT match to ARs

# Match Status by State and Survey Question

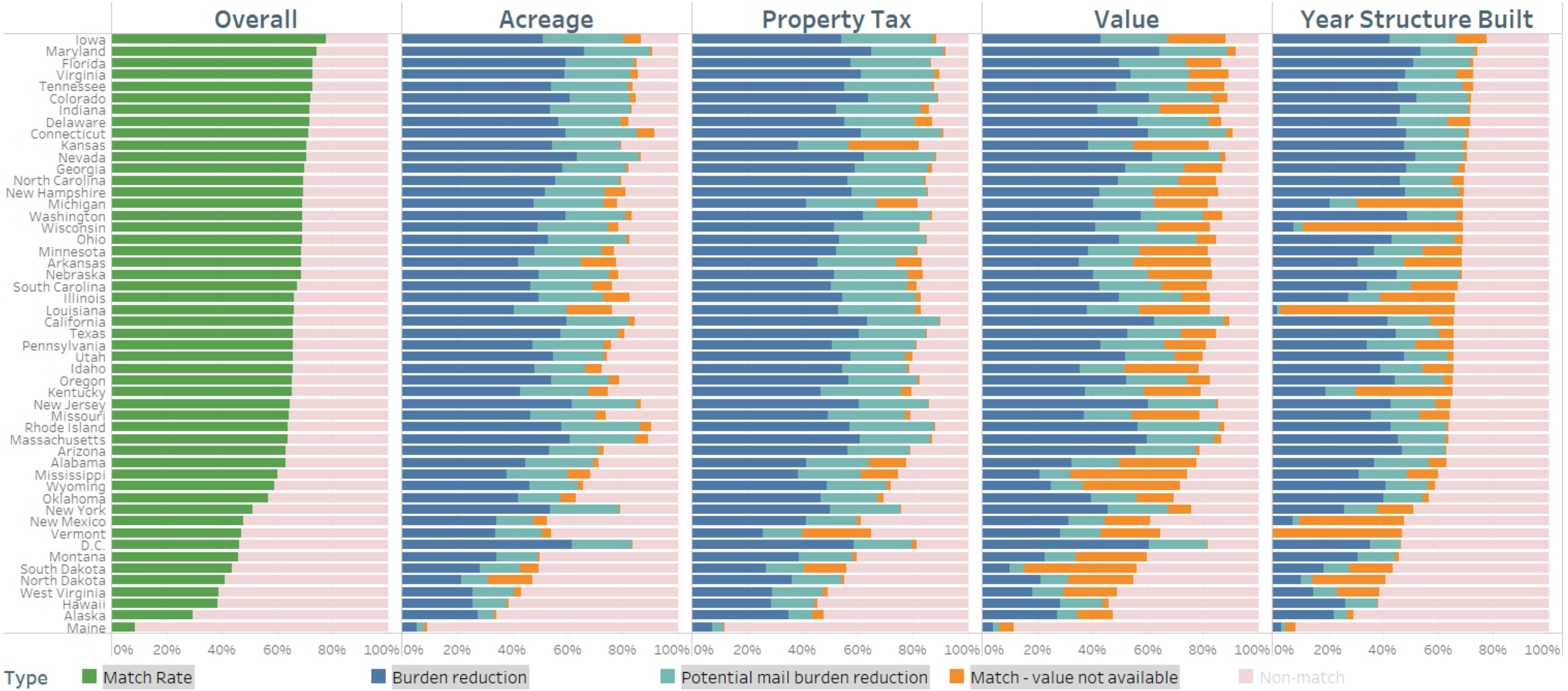


# Type of Match

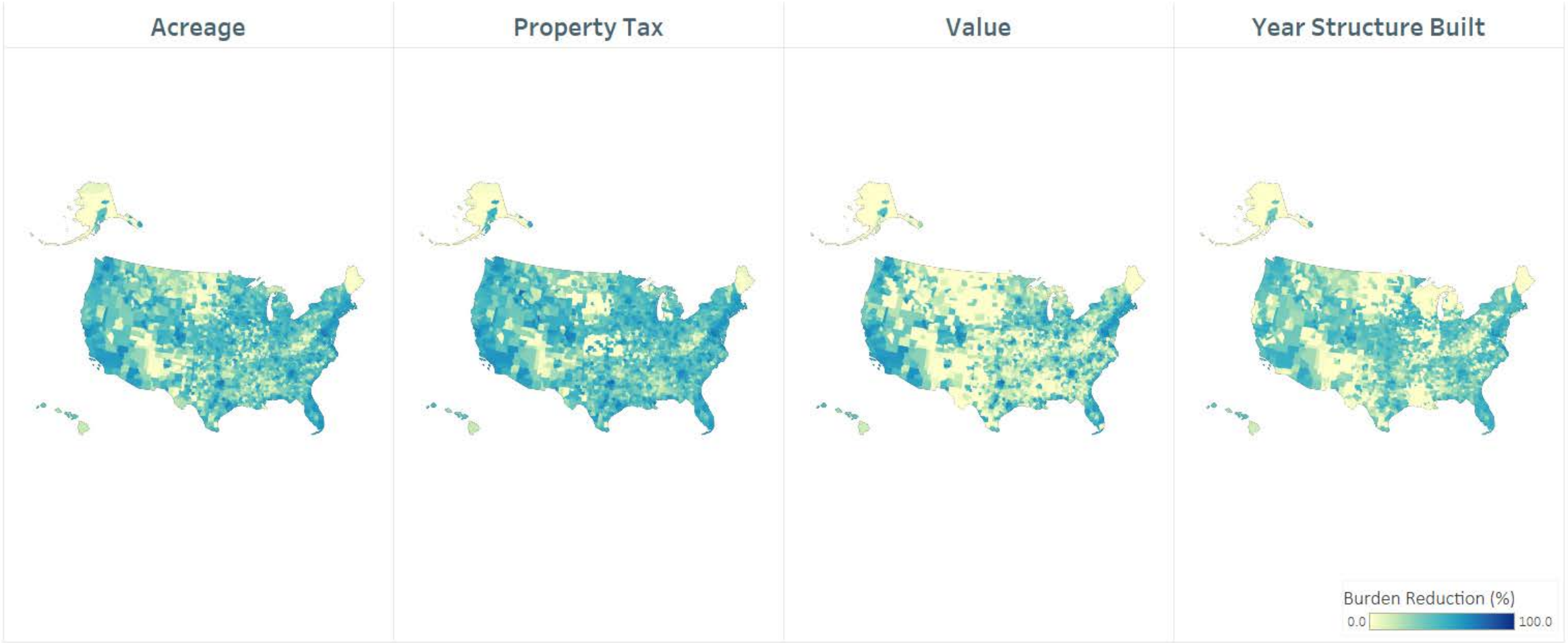
**Match** = ACS responding household with a AR value for at least one of the 4 tested items

- **Burden reduction** = match that would not be asked question in our adaptive design
- **Potential mail burden reduction** = match that would not be asked question if we included the mail mode in our adaptive design
- **Match - value not available** = match but a value for question is not available

# Match Status by State and Survey Question



# Burden Reduction by County and Survey Question – Geographic View




*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.*

# Conclusions

- Potential to significantly reduce respondent burden for these housing topics
- Significant improvement in item allocation rates
- Challenges:
  - Reliance on outside vendor and proprietary models
  - Differences in coverage and availability of data
  - Often a time lag between survey year and AR
  - Issues with matching administrative records to census/survey records
  - Using AR complicates disclosure avoidance rules
  - Break in time series
  - Unintended impacts on other survey items

# The Promise of Administrative Records

- Leveraging existing data sources through linked approaches will be an important component of demographic research in the coming years.
- The field of survey research is shifting and we must do what we can to leverage other data sources to enhance, supplement, or in some cases replace what we gain from surveys.
- The Census Bureau is engaging the use of administrative records at all stages of the survey life cycle.

# Thank You!

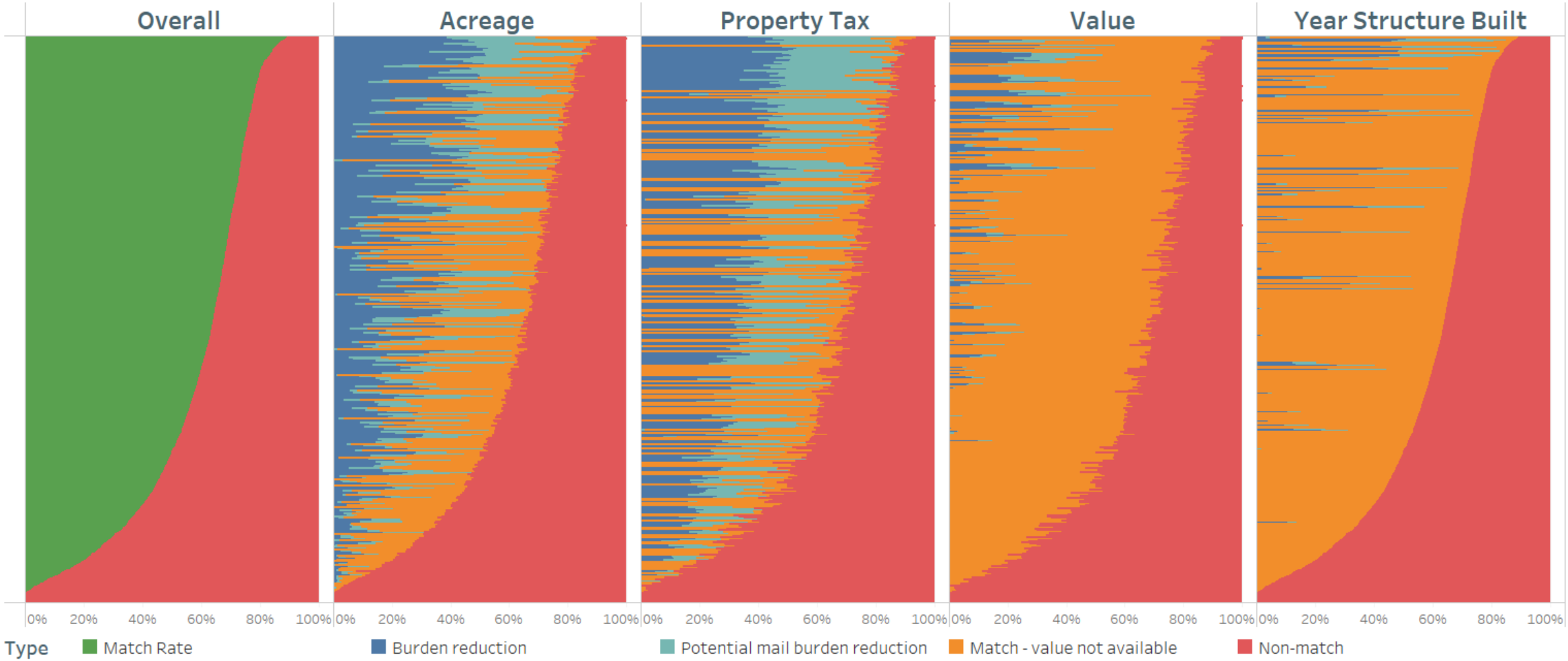
Jennifer.M.Ortman@census.gov

301-763-5750

# Supplemental Material

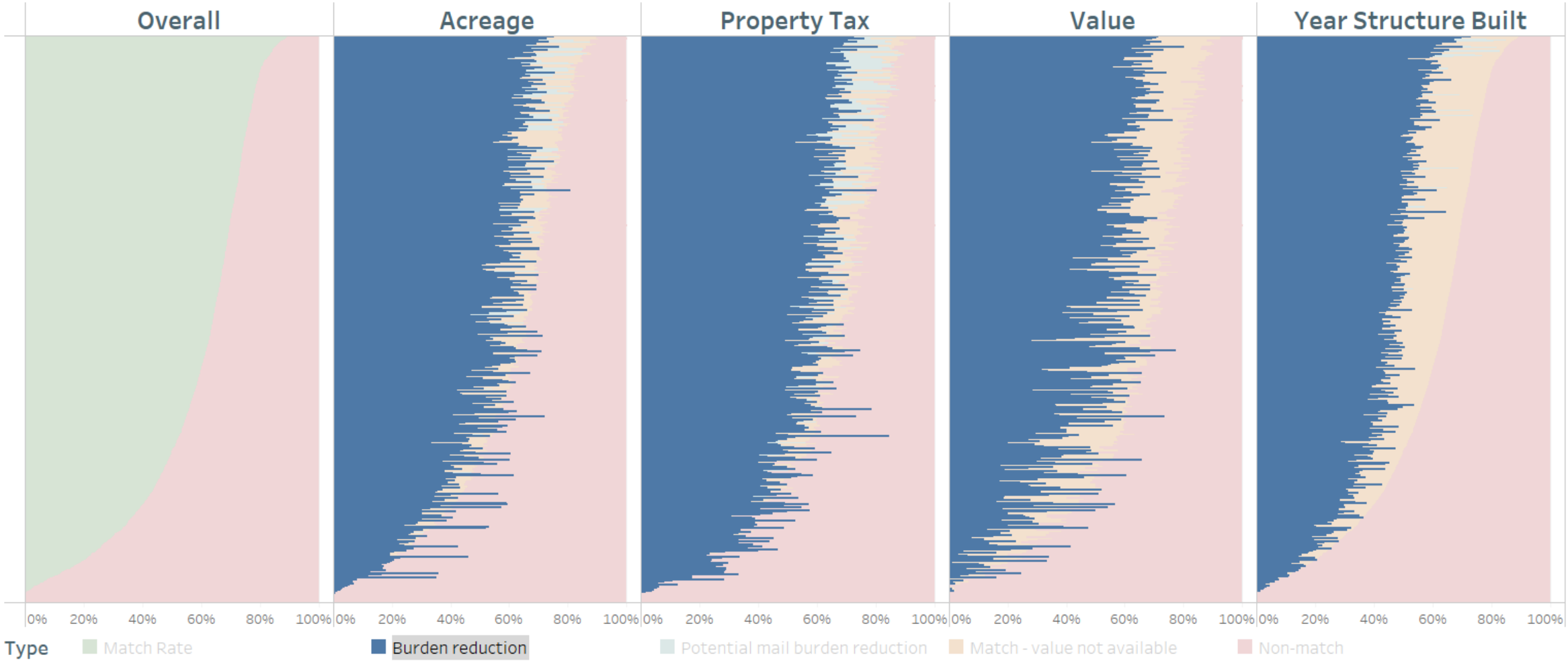
# Match Status by County and Survey Question

Sorted by percentage of household with any admin record matches



# Burden Reduction by County and Survey Question

Sorted by percentage of household with any admin record matches




United States Census Bureau

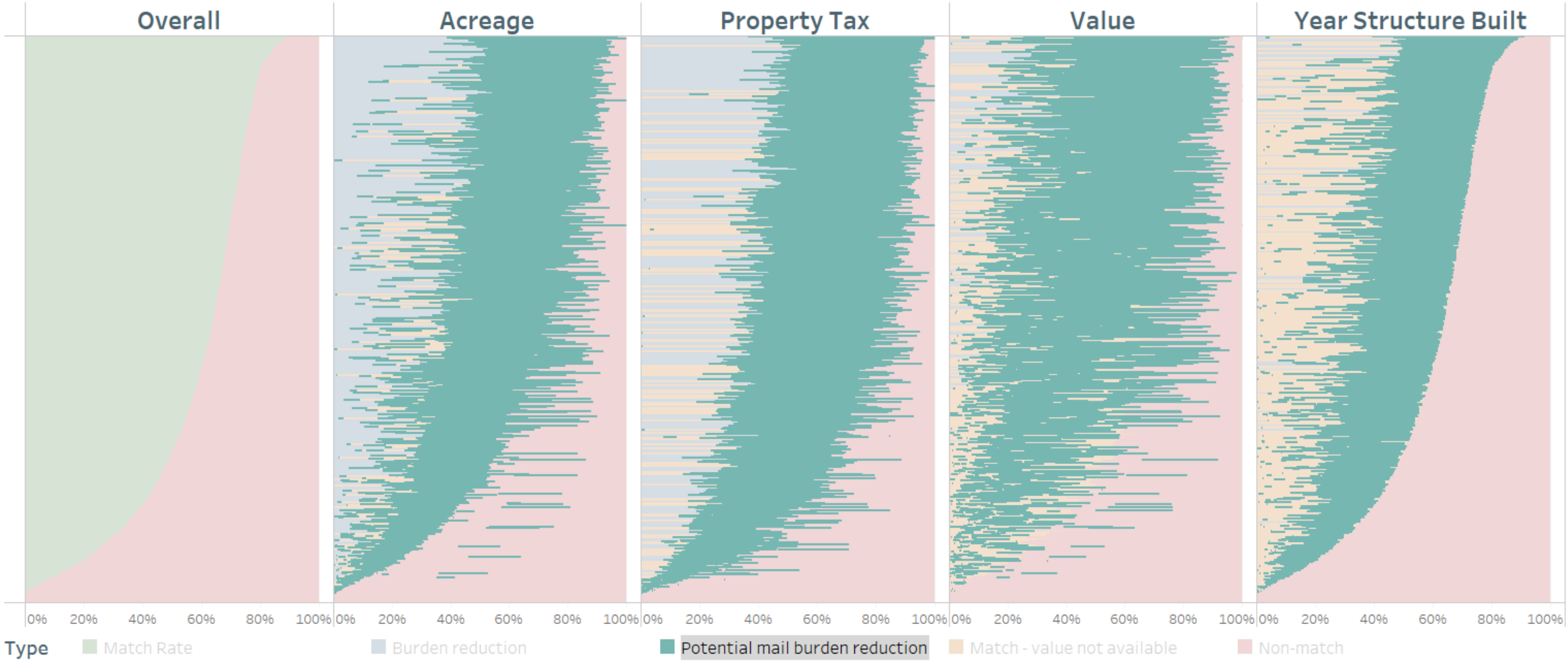
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

**Overall Burden Reduction by Survey Question:**

**Acreage - 52%  Real Estate Tax - 54%  Property Value - 47%  Year Built - 38%**

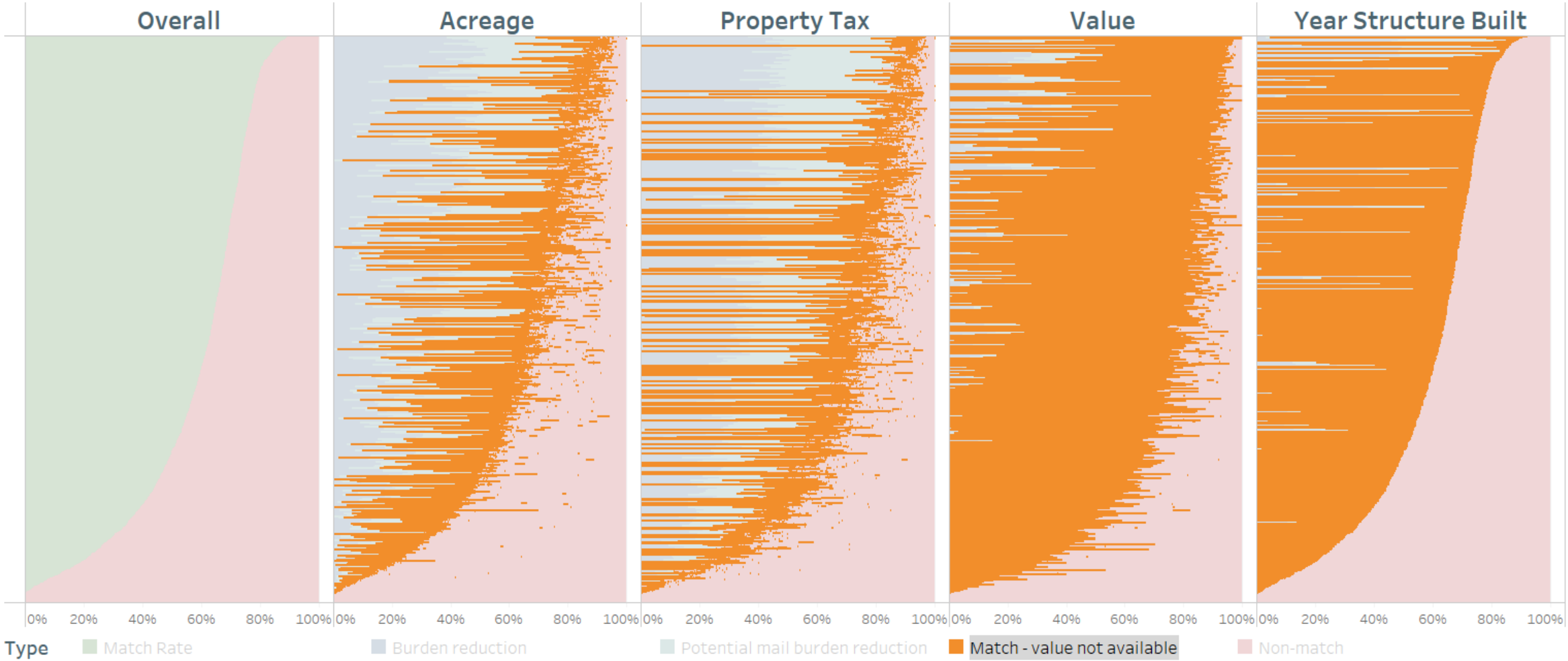
# Potential Mail Burden Reduction by County and Survey Question

Sorted by percentage of household with any admin record matches



# Match to AR but Question Item Not Available by County and Survey Question

Sorted by percentage of household with any admin record matches



# Guiding Principles

| | |
|---|---|
| Authorization | • Do we have a formal agreement (e.g., contract or interagency agreement) to obtain and use an administrative data? |
| Availability | • Are the data available for every year? |
| Conceptual Alignment | • Do the administrative data correspond to the concept the ACS currently intends to measure? |
| Coverage | • How comprehensive is the coverage of the administrative data with respect to geographies and population subgroups? |
| Data Source | • Do the administrative data come from a trusted and respected source, above reproach and conflict free? |
| Disclosure Avoidance | • Does use of the administrative data preclude the Census Bureau from ensuring disclosure avoidance of personally identifiable information? |

# Guiding Principles

| | |
|---|---|
| Impacts on Estimates | • To what extent does the administrative data source diverge from survey response? Do the differences carry over to other variables (e.g., through editing and imputation)? |
| Intended Use | • How will the administrative data be used (e.g., editing and imputation, substitution, blended data product)? |
| Population Universe | • Are the administrative data intended for use to measure something for the total U.S. population or a population subgroup (e.g., condo owners)? |
| Quality | • What is sufficient data quality for the published estimates? Do the administrative data meet these quality requirements? |
| Reliability | • Are the administrative data available and consistent over time? |
| Temporal Alignment | • Do the administrative data correspond to the time period referenced in the ACS? |