2019 Multiple Imputation and Hot Deck Methods in the American Housing Survey

Stephen Ash, Kathy Zha, Sean Dalby, and Gregory Mulley
Demographic Statistical Methods Division

July 30, 2019

Presenter: Sean Dalby
Email: sean.dalby@census.gov

Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Topics

1. The American Housing Survey (AHS): Introduction and Details
2. Imputation: Hot Deck and Multiple Imputation (MI) in AHS
3. Results: MI Diagnostics and Comparison to Hot Deck
4. Next Steps and Future Research

1. The American Housing Survey: Content

- Nationally representative longitudinal sample of all housing units within the US – survey returns to selected housing units every 2 years

- Complex survey design – 2 stage sampling using stratified and systematic sampling methods – 85,931 units sampled in 2019

- Detailed questionnaire inquiring about demographics, tenure status, income, housing costs, number of rooms, type of plumbing, heating, etc.

- Started in 1973 – sponsored by Department of Housing and Urban Development (HUD)

1. The American Housing Survey: Missing Data

- This presentation deals solely with item nonresponse
- Two types of missing data in AHS: unit and item nonresponse
- Unit nonresponse occurs when no information at all is acquired from a respondent at a housing unit - handled with weighting adjustment
- Item nonresponse occurs when a respondent selectively refuses to answer some questions, but otherwise responds to the survey

1. The American Housing Survey: Imputation Difficulties

- Several sources of difficulty when it comes to imputing AHS data:
  - Multiple kinds of variables: continuous, binomial, count, categorical, etc.
  - Binning problems and other distributional irregularities
  - Structural zeros: dependencies / skip-patterns

- Examples of structural zeros:
  - *Dependencies:* Initial question about housing unit type dictates values for future questions about housing unit type
  - *Skip patterns:* Only home owners are asked about mortgages – cannot impute a mortgage for renters

2. Imputation: AHS Application of Hot Deck

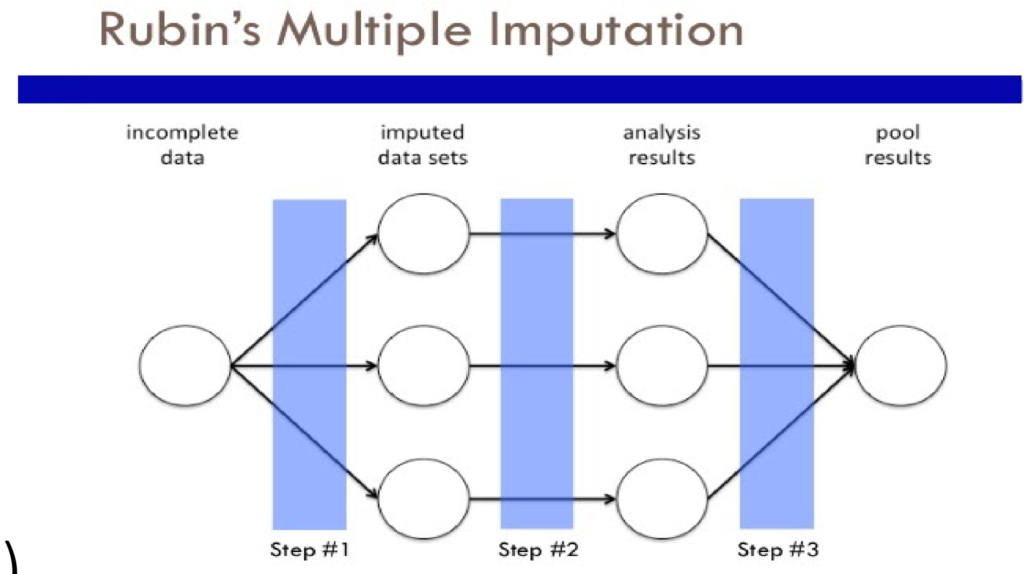
Imputed Variable	Variable Description
NUNIT2	What is the structure type of this unit?
TYPE	Is the unit a house, apartment, mobile home or another type of residence?
NUNITS	How many units are in this building?

Source: U.S. Census Bureau, 2017 American Housing Survey

Number of floors	Tenure/Vacancy Status	Result
1 Floor	Owner Occupied or Vacant for Sale or Off Market	Imputation cell 1
	Renter Occupied or Vacant for Rent	Imputation cell 2
2 – 3 Floors	Owner Occupied or Vacant for Sale or Off Market	Imputation cell 3
	Renter Occupied or Vacant for Rent	Imputation cell 4
4 Floors	Owner Occupied	Imputation cell 5
	Renter or Vacant or Usual Residence Elsewhere (URE)	Imputation cell 6
5 or more Floors	All Housing Units	Imputation cell 7
Unknown number of floors	Owner Occupied or Vacant for Sale or Off Market	Imputation cell 8
	Renter Occupied or Vacant for Rent	Imputation cell 9

2. Imputation: Theory Behind Multiple Imputation

- Impute multiple values for each missing observation - captures uncertainty about the right value to impute
- Analyze and combine results from these multiple datasets using standard procedures (Rubin 1987)
- Using Fully Conditional Specification (FCS) (van Buuren 2019) / Sequential Regressions (Raghunathan 2016)



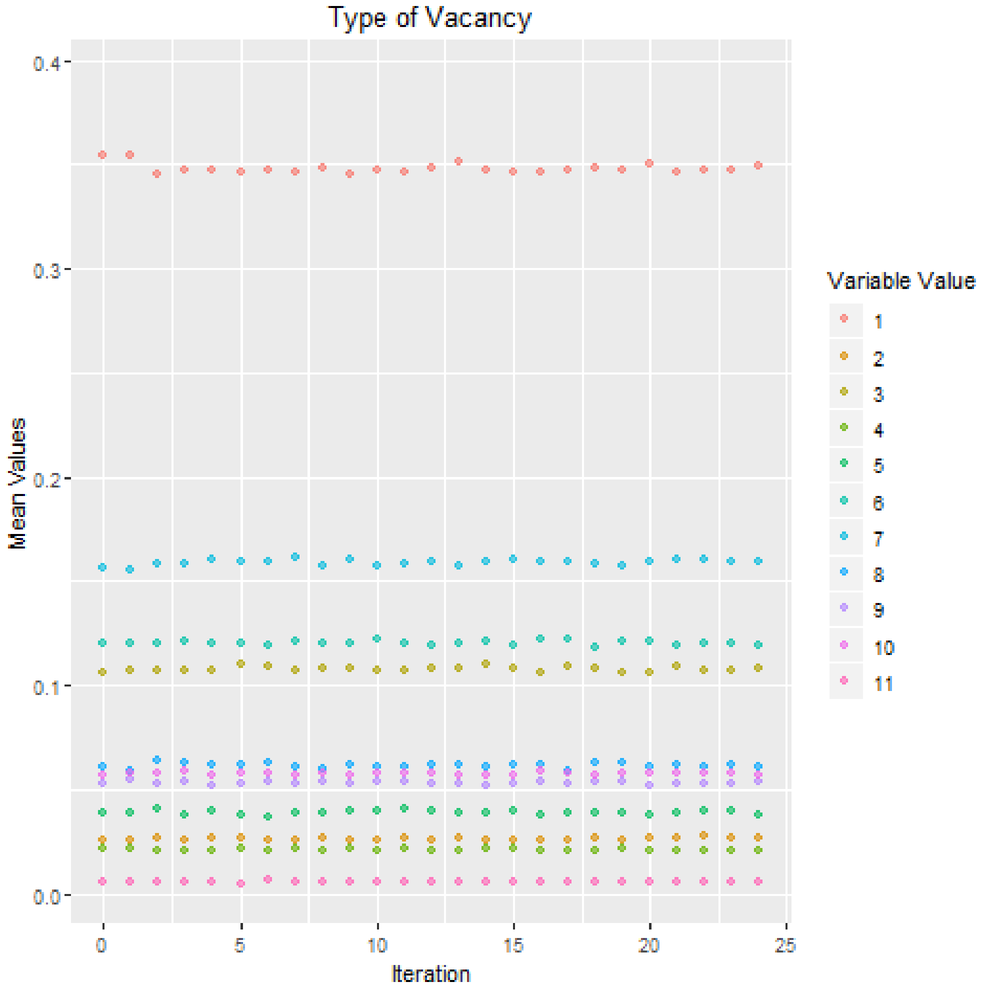
2. Imputation: Benefits / Tradeoffs

- Well developed, broad modeling can produce:
  - unbiased point estimates and variances
  - preserves covariances
- Flexible modeling under FCS – handles structural zeros as well
- Measures variance of imputation itself
- Model contingent
- Computationally expensive

2. Imputation: Important Considerations for MI

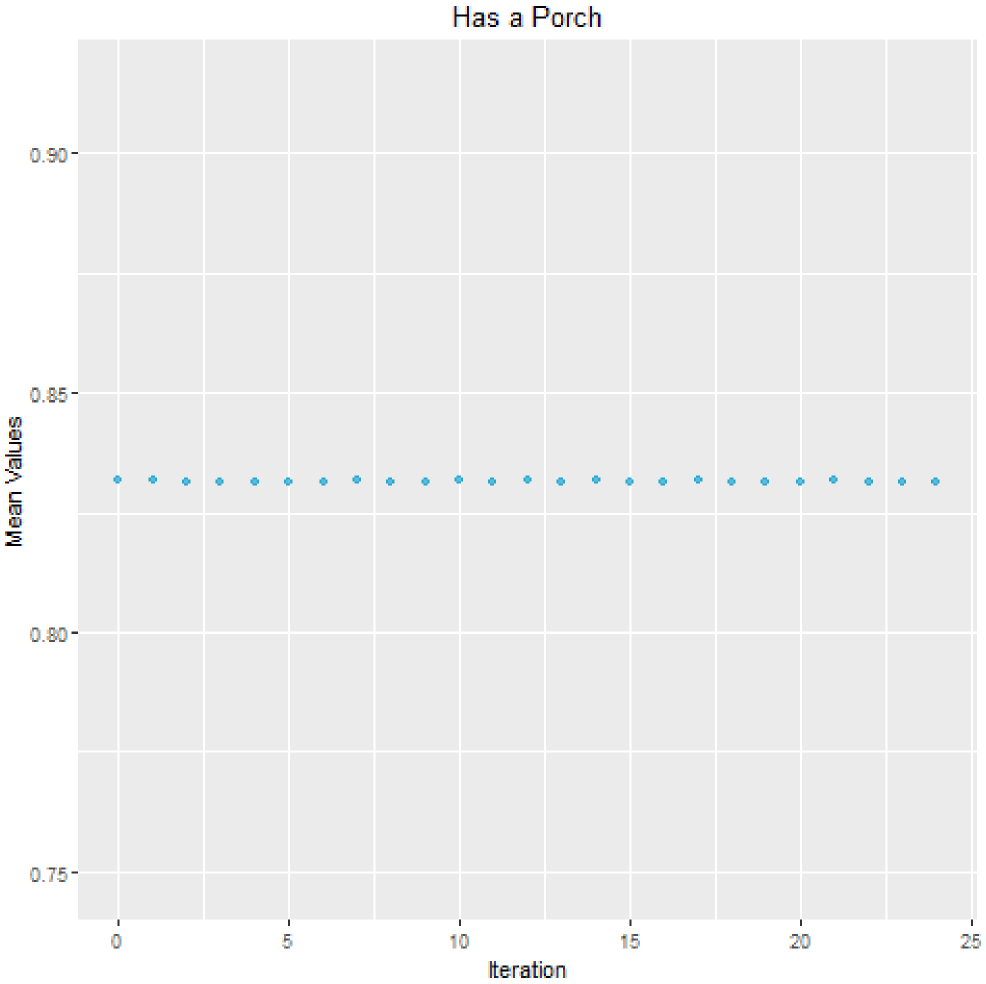
- How many datasets to choose? (M = 5 is standard)
- Should missingness be modeled? (Ignorability)
- Scope of variables used in modeling? (Congeniality)
- Do conditional distributions stem from one joint distribution? (Compatibility)
- Choice of imputation method?
  - Chose Bayesian FCS, but others – like predictive mean matching – can work too
- Non-linearity or non-parametric methods available?
  - Used non-linear Principal Component Analysis as a variable reduction method

3. Results: Convergence of Means



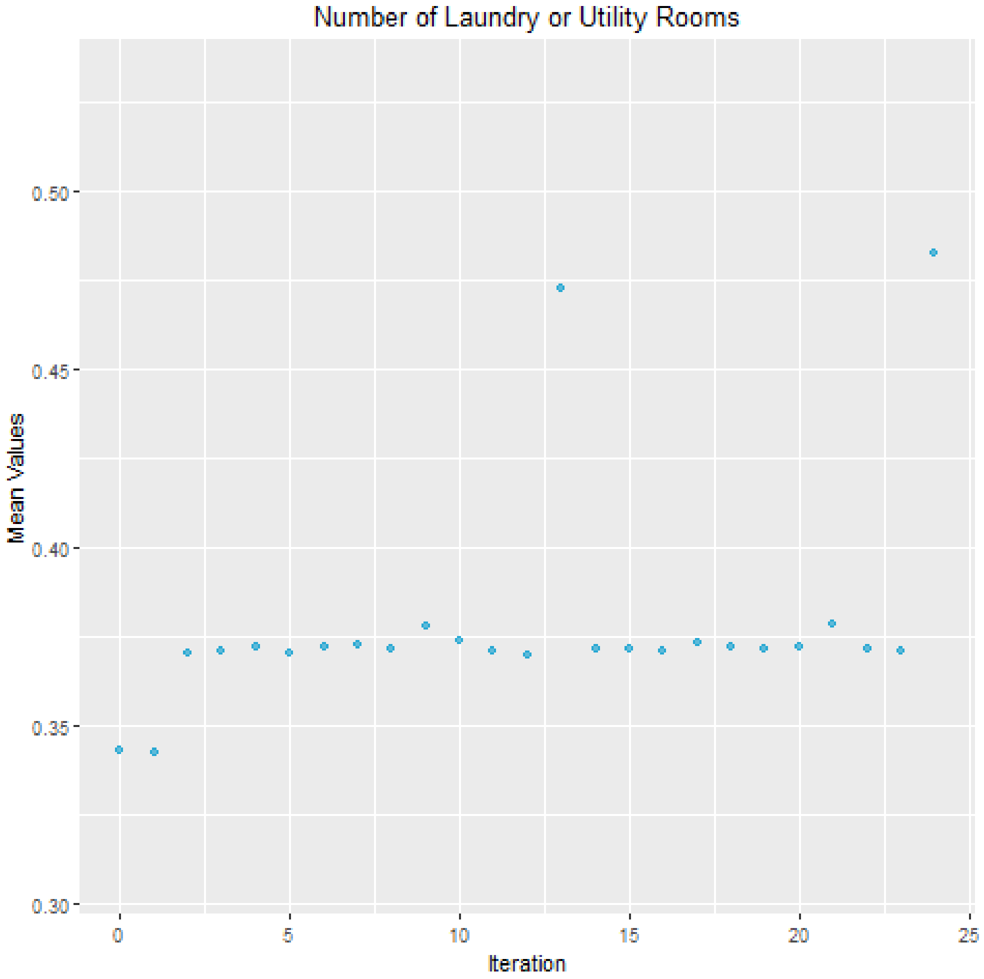
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Results: Convergence of Means



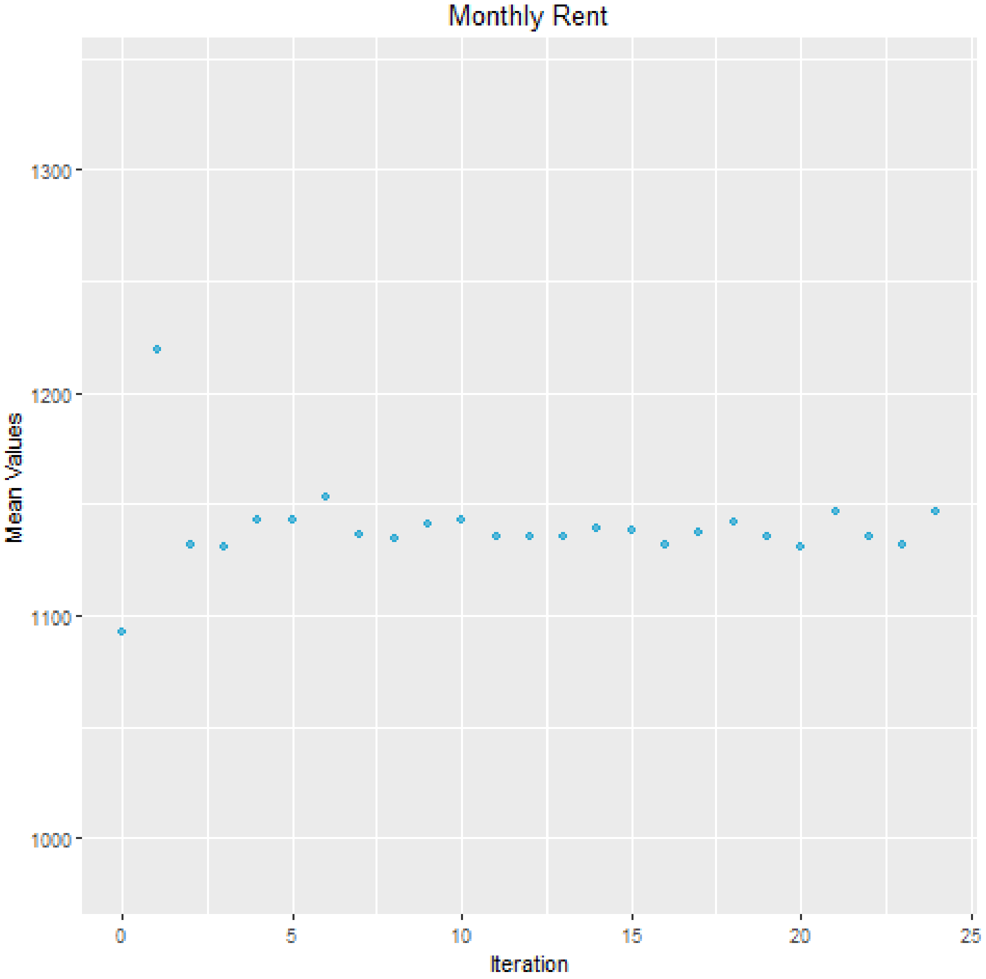
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Results: Convergence of Means



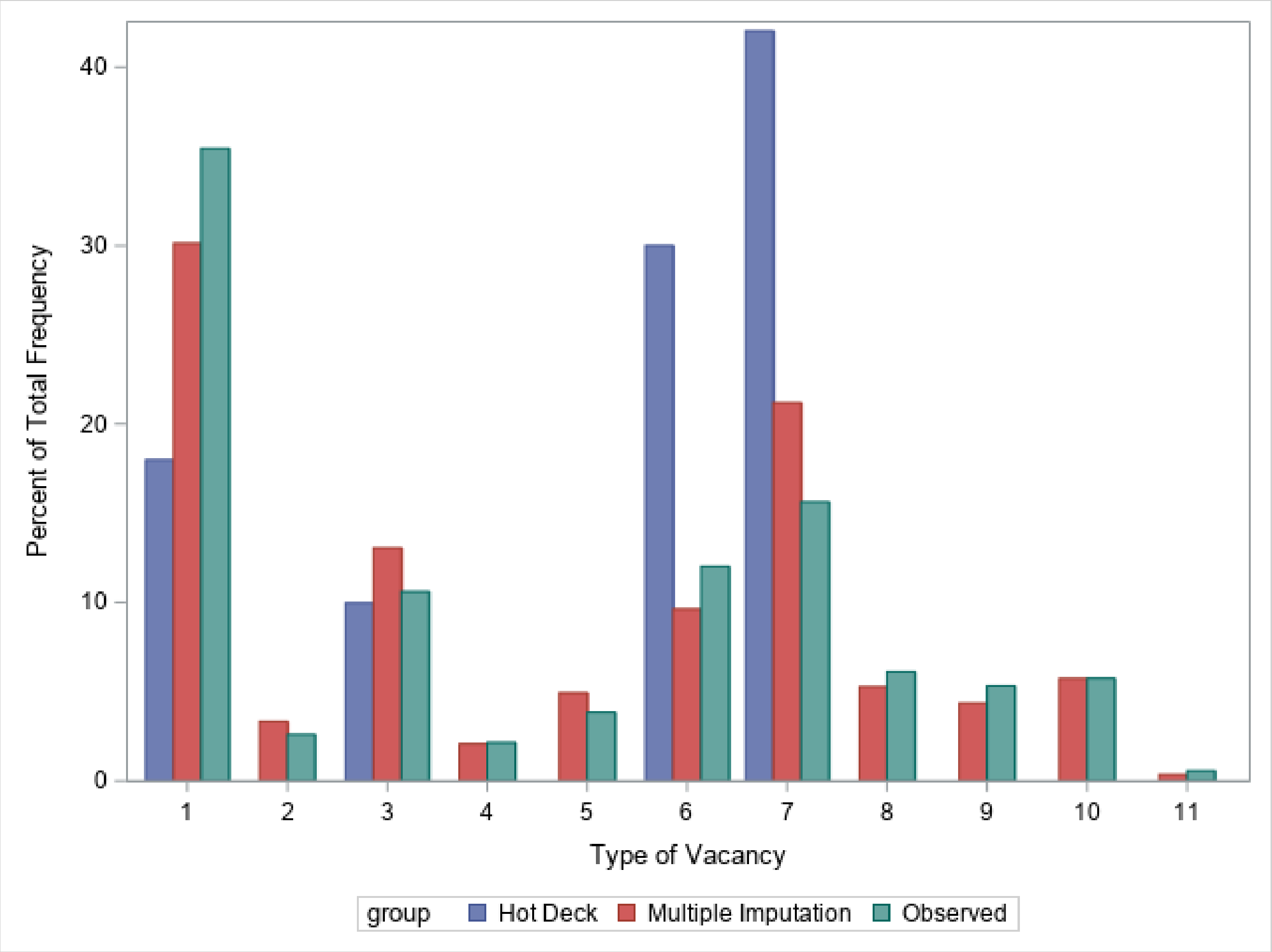
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Results: Convergence of Means



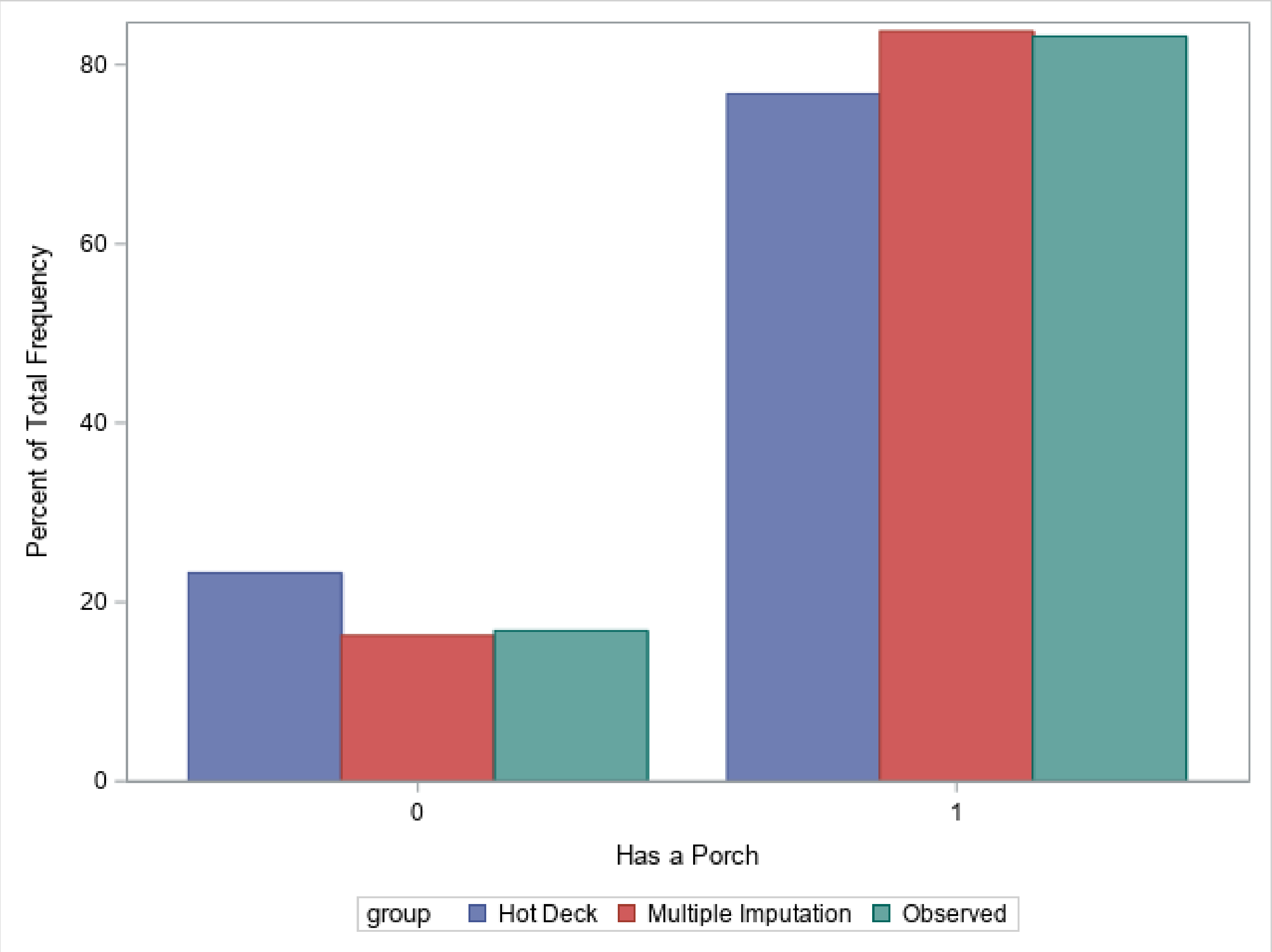
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Hot Deck vs. Multiple Imputation: Real Data



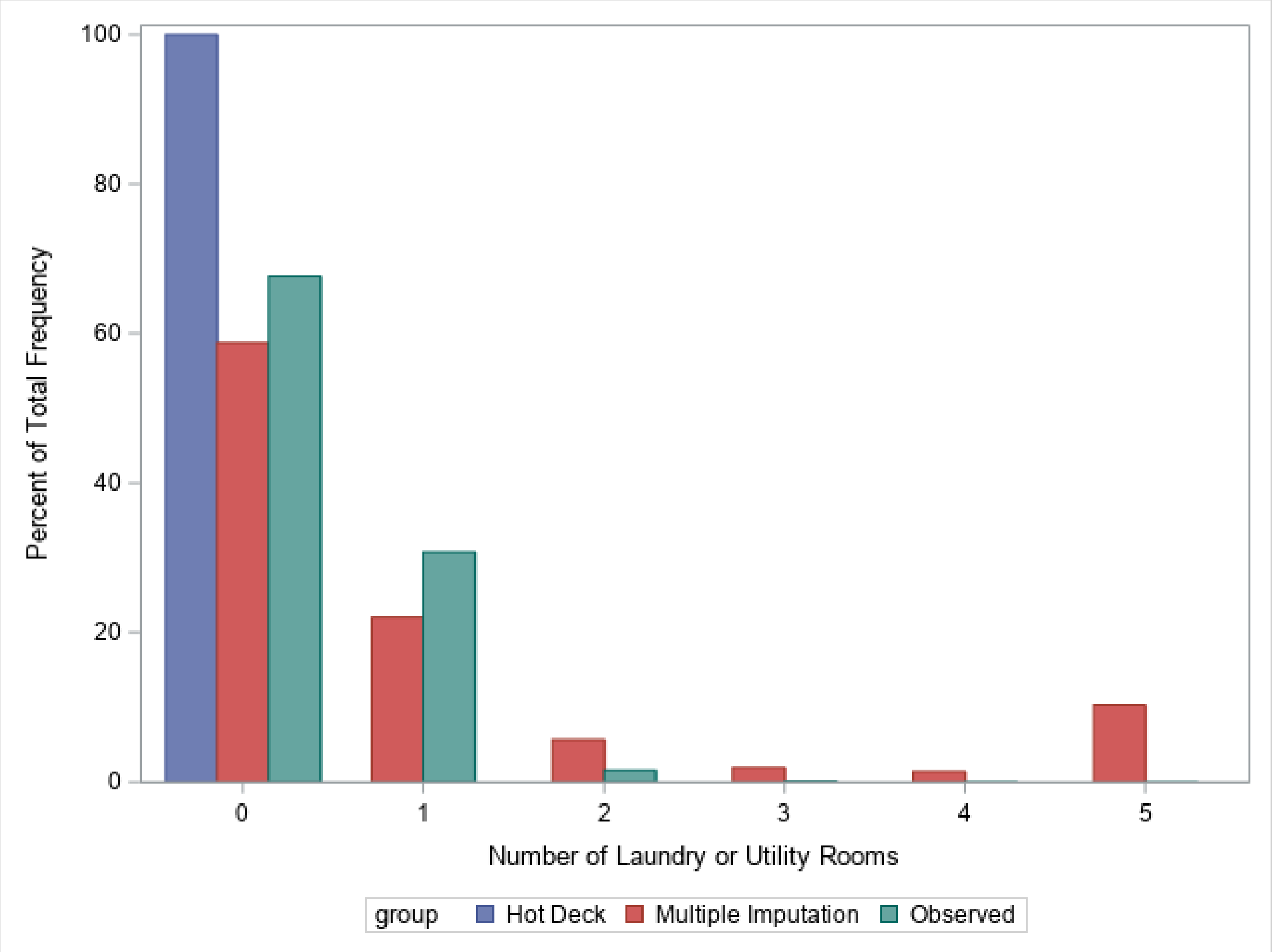
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Hot Deck vs. Multiple Imputation: Real Data



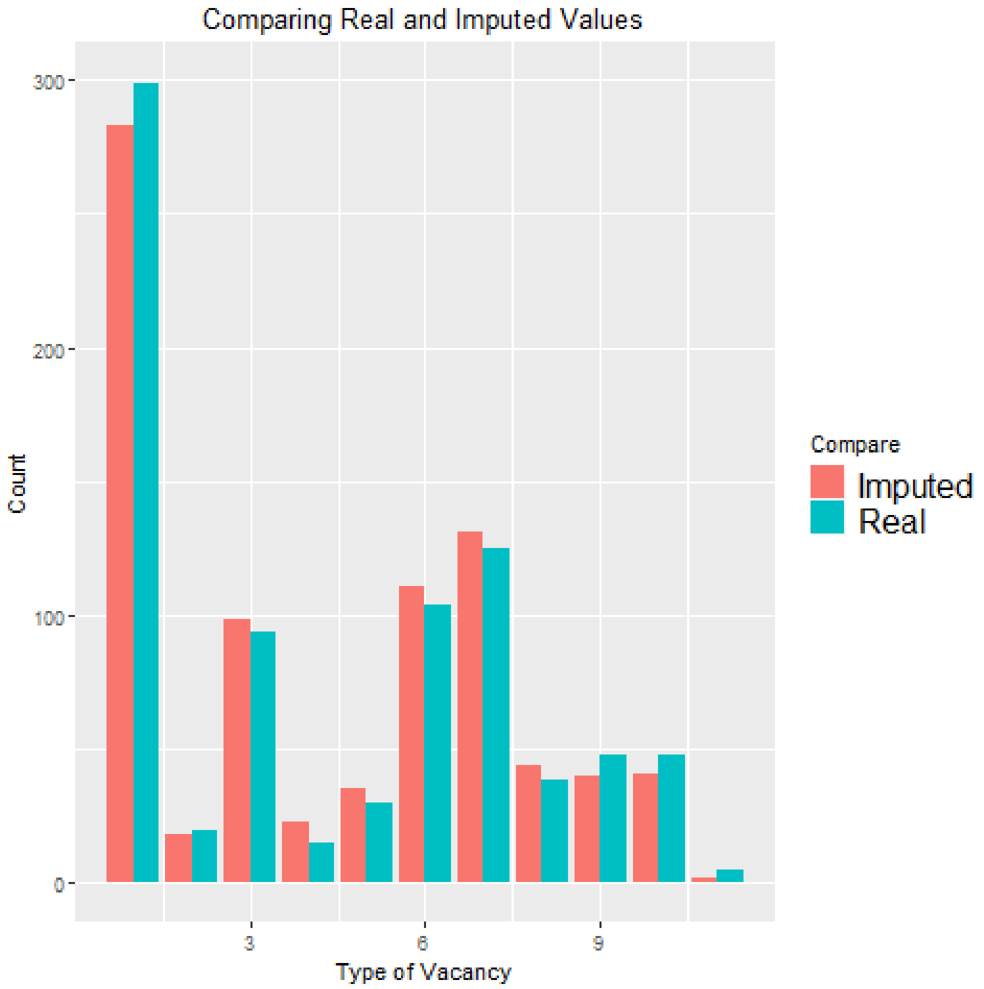
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Hot Deck vs. Multiple Imputation: Real Data



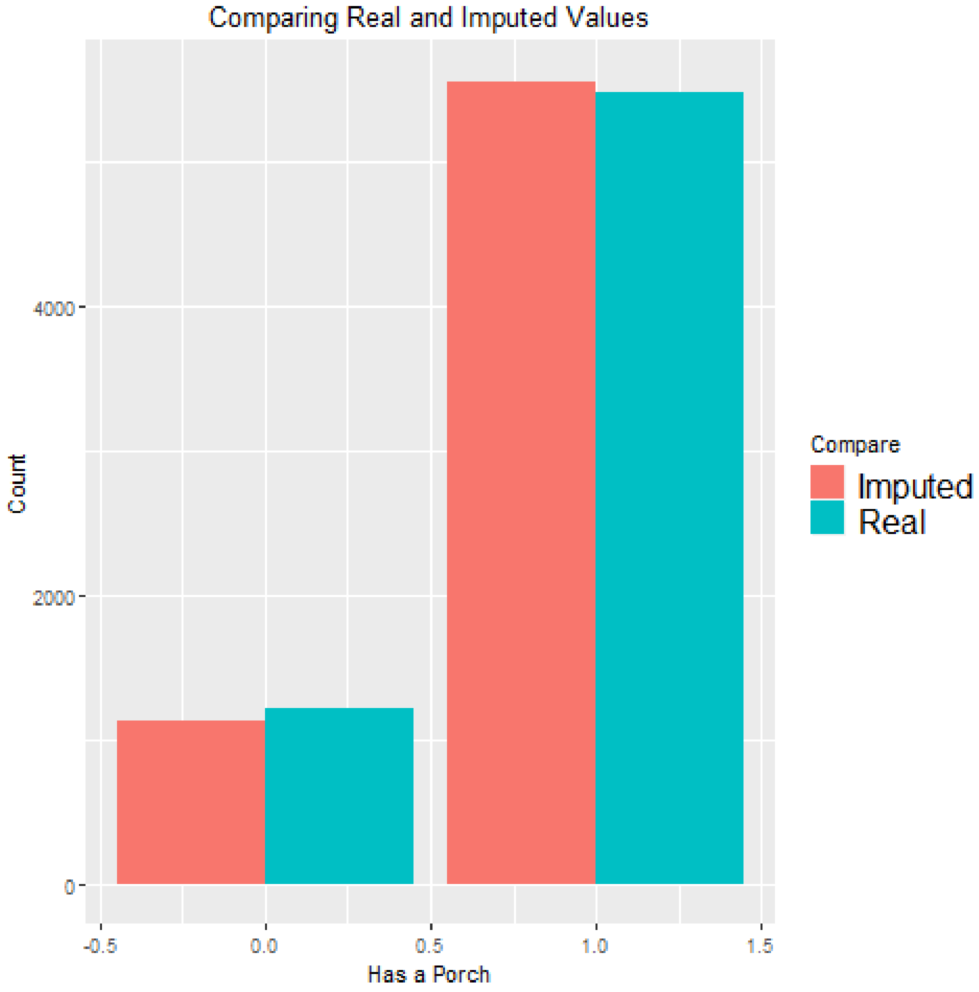
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Multiple Imputation: Validation Data



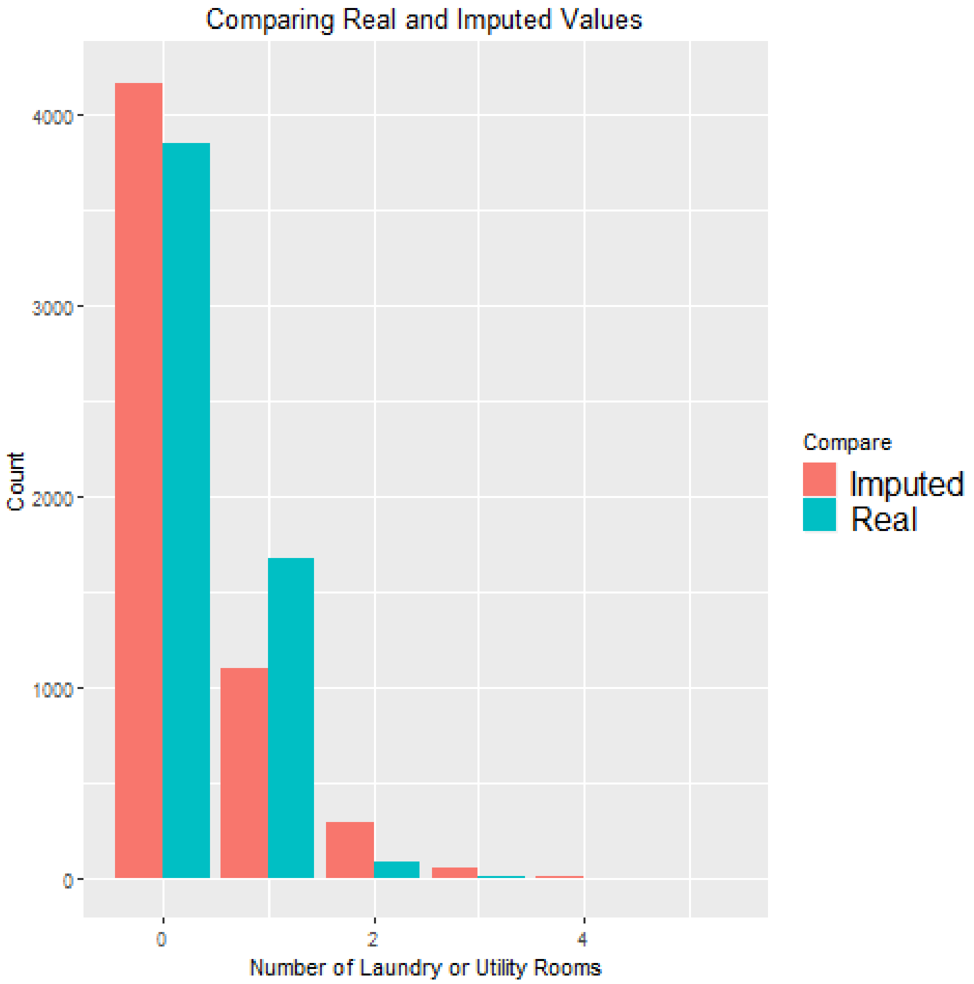
Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Multiple Imputation: Validation Data



Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Multiple Imputation: Validation Data



Source: U.S. Census Bureau, 2017 American Housing Survey

3. Observed vs. Hot Deck – Difference in Correlations

Variable	RENT	BUILT	VALUE	CSTMNT	UNITSF	LOT	GARAGE	ANCHOR	NOSTEP	GUTREHB	HMRACCE	HMRENEF	HMRSALE	ACCESS	ACCESSB	COMPLEX	PORCH
RENT	0.000	0.010			0.013	0.063	0.010	0.002	0.005						0.006	0.001	0.000
BUILT	0.010	0.000	0.002	0.002	0.000	0.042	0.001	0.019	0.002	0.001	0.001	0.000	0.001		0.000	0.004	0.001
VALUE		0.002	0.000	0.019	0.092	0.022	0.028	0.050	0.001	0.001	0.005	0.000	0.001		0.015	0.033	0.009
CSTMNT		0.002	0.019	0.000	0.000	0.018	0.000	0.004	0.000	0.000	0.001	0.001	0.000		0.006	0.007	0.000
UNITSF	0.013	0.000	0.092	0.000	0.000	0.025	0.000	0.018	0.000	0.000	0.001	0.001	0.001		0.003	0.000	0.000
LOT	0.063	0.042	0.022	0.018	0.025	0.000	0.024	0.038	0.047	0.023	0.018	0.007	0.014				0.015
GARAGE	0.010	0.001	0.028	0.000	0.000	0.024	0.000	0.009	0.000	0.000	0.001	0.001	0.000		0.010	0.003	0.001
ANCHOR	0.002	0.019	0.050	0.004	0.018	0.038	0.009	0.000	0.002	0.009	0.003	0.010	0.004				0.016
NOSTEP	0.005	0.002	0.001	0.000	0.000	0.047	0.000	0.002	0.000	0.000	0.001	0.001	0.000		0.007	0.001	0.000
GUTREHB		0.001	0.001	0.000	0.000	0.023	0.000	0.009	0.000	0.000	0.002	0.002	0.001		0.007	0.001	0.000
HMRACCESS		0.001	0.005	0.001	0.001	0.018	0.001	0.003	0.001	0.002	0.000	0.003	0.001		0.005	0.009	0.002
HMRENEFF		0.000	0.000	0.001	0.001	0.007	0.001	0.010	0.001	0.002	0.003	0.000	0.000		0.001	0.009	0.001
HMRSALE		0.001	0.001	0.000	0.001	0.014	0.000	0.004	0.000	0.001	0.001	0.000	0.000		0.004	0.002	0.001
ACCESS																	
ACCESSB	0.006	0.000	0.015	0.006	0.003		0.010		0.007	0.007	0.005	0.001	0.004		0.000	0.001	0.007
COMPLEX	0.001	0.004	0.033	0.007	0.000		0.003		0.001	0.001	0.009	0.009	0.002		0.001	0.000	0.004
PORCH	0.000	0.001	0.009	0.000	0.000	0.015	0.001	0.016	0.000	0.000	0.002	0.001	0.001		0.007	0.004	0.000

Source: U.S. Census Bureau, 2017 American Housing Survey

0.00 0.03 0.06

3. Observed vs. MI – Difference in Correlations

Variable	RENT	BUILT	VALUE	CSTMNT	UNITSF	LOT	GARAGE	ANCHOR	NOSTEP	GUTREHB	HMRACCE	HMRENEF	HMRSALE	ACCESS	ACCESSB	COMPLEX	PORCH
RENT	0.000	0.017			0.008	0.013	0.001	0.007	0.002					0.006	0.019	0.004	0.008
BUILT	0.017	0.000	0.019	0.008	0.003	0.001	0.022	0.008	0.011	0.002	0.013	0.039	0.006	0.001	0.026	0.022	0.013
VALUE		0.019	0.000	0.057	0.057	0.001	0.014	0.035	0.007	0.013	0.014	0.007	0.007	0.001	0.031	0.057	0.021
CSTMNT		0.008	0.057	0.000	0.026	0.005	0.005	0.004	0.003	0.016	0.014	0.028	0.011	0.000	0.017	0.014	0.005
UNITSF	0.008	0.003	0.057	0.026	0.000	0.000	0.013	0.032	0.000	0.001	0.003	0.004	0.000	0.002	0.003	0.005	0.014
LOT	0.013	0.001	0.001	0.005	0.000	0.000	0.008		0.002	0.005	0.000	0.001	0.003	0.000			0.002
GARAGE	0.001	0.022	0.014	0.005	0.013	0.008	0.000	0.013	0.000	0.000	0.004	0.009	0.003	0.000	0.002	0.001	0.000
ANCHOR	0.007	0.008	0.035	0.004	0.032		0.013	0.000	0.002	0.008	0.004	0.001	0.017				0.023
NOSTEP	0.002	0.011	0.007	0.003	0.000	0.002	0.000	0.002	0.000	0.001	0.001	0.009	0.001	0.000	0.003	0.001	0.000
GUTREHB		0.002	0.013	0.016	0.001	0.005	0.000	0.008	0.001	0.000	0.020	0.038	0.002	0.000	0.010	0.002	0.000
HMRACCESS		0.013	0.014	0.014	0.003	0.000	0.004	0.004	0.001	0.020	0.000	0.005	0.011	0.004	0.031	0.016	0.012
HMRENEFF		0.039	0.007	0.028	0.004	0.001	0.009	0.001	0.009	0.038	0.005	0.000	0.004	0.003	0.004	0.043	0.001
HMRSALE		0.006	0.007	0.011	0.000	0.003	0.003	0.017	0.001	0.002	0.011	0.004	0.000	0.006	0.069	0.002	0.005
ACCESS	0.006	0.001	0.001	0.000	0.002	0.000	0.000		0.000	0.000	0.004	0.003	0.006	0.000	0.005	0.001	0.000
ACCESSB	0.019	0.026	0.031	0.017	0.003		0.002		0.003	0.010	0.031	0.004	0.069	0.005	0.000	0.002	0.003
COMPLEX	0.004	0.022	0.057	0.014	0.005		0.001		0.001	0.002	0.016	0.043	0.002	0.001	0.002	0.000	0.000
PORCH	0.008	0.013	0.021	0.005	0.014	0.002	0.000	0.023	0.000	0.000	0.012	0.001	0.005	0.000	0.003	0.000	0.000

Source: U.S. Census Bureau, 2017 American Housing Survey

0.00 0.03 0.06

4. Next Steps and Future Research

Additional Testing

- Extend the scope
- Model refinements
- More diagnostics
- Measure variance introduced by imputation

Implementation in 2019 AHS

- Scope – modules to be included
- Timing to fit in production schedule

Thanks!

Presenter: Sean Dalby

Email: sean.dalby@census.gov