



Adaptable and Efficient Generation of RTI's 2017 Synthetic Population



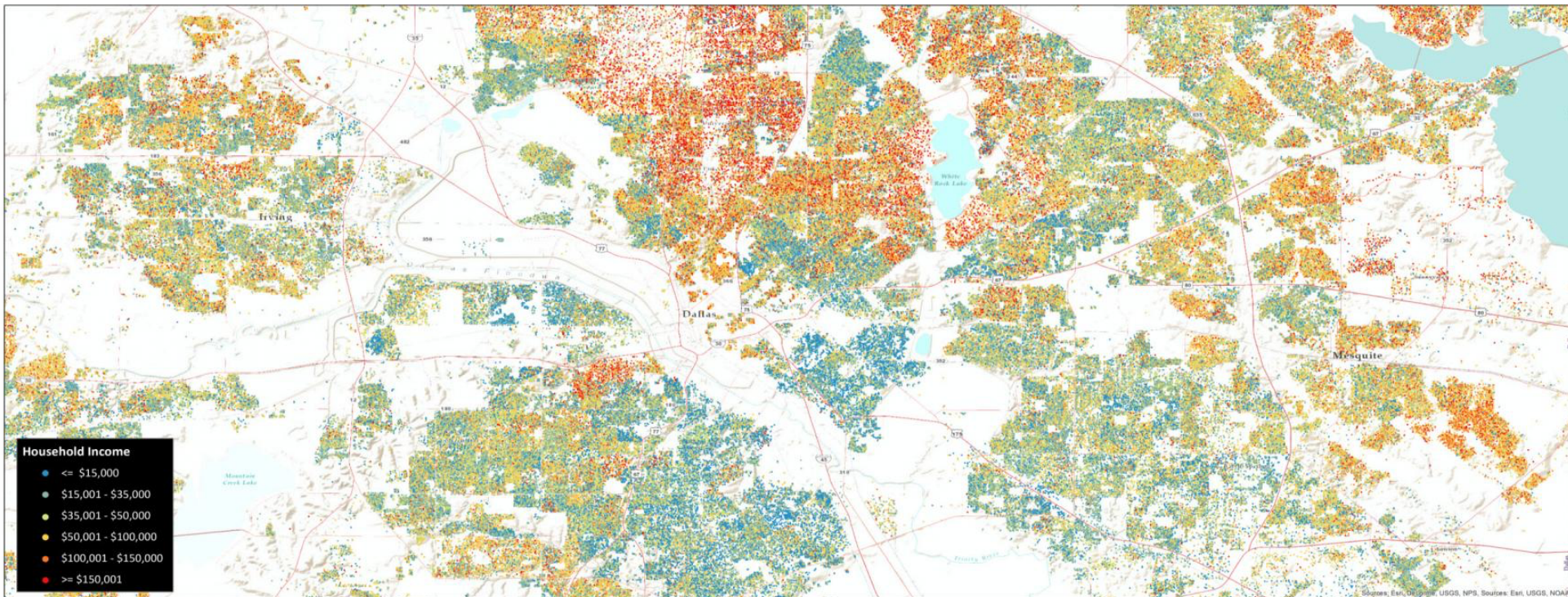
RTI's Center for Data Science and Geospatial Science and Technology Program

Ishan Shah




Overview of RTI's Synthetic Population

- **Goal:** Generate detailed data for every household and person in the US, as accurate as possible, while maintaining anonymity
- Previous uses of 2010, 2013 populations
 - Locating at-risk obesity populations
 - Estimating underage drinking population
 - Modeling spread of Infectious disease
 - Predicting hot real estate neighborhoods



2013-17 American Community Survey (ACS) Data

Detailed survey sent to ~3.5M
households/year



```
graph TD; A[2013-17 American Community Survey (ACS) Data  
Detailed survey sent to ~3.5M households/year] --> B[Aggregates - 5-year estimated counts at block group (600-3000 people) level]; A --> C[Public Use Microdata Sample (PUMS) - Anonymized 5% sample of responses at PUMA (~100,000 people) level];
```

Aggregates - 5-year
estimated counts at block
group (600-3000 people)
level

**Public Use Microdata
Sample (PUMS)** -
Anonymized 5% sample of
responses at PUMA
(~100,000 people) level

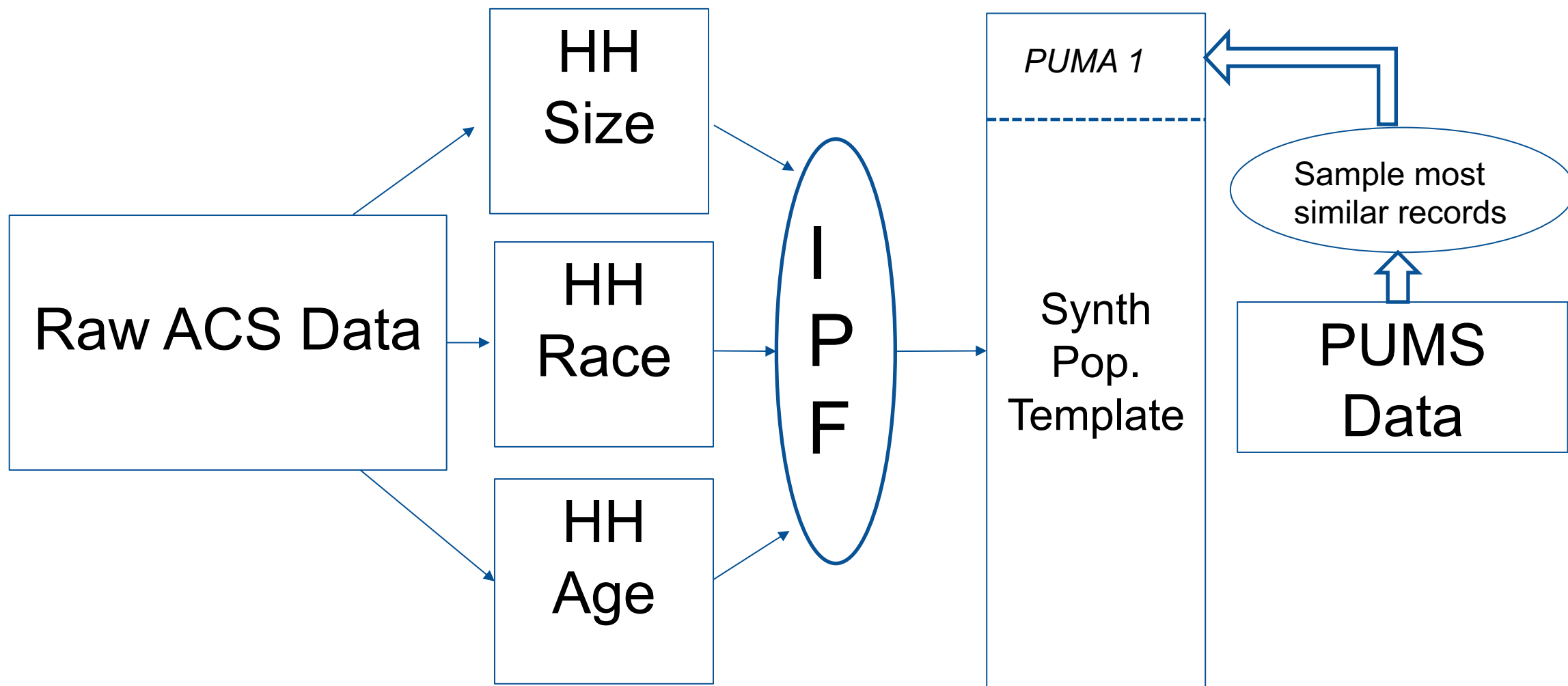
The Process

1. Download lots of data

2. Create Marginal Tables

3. Use Iterative Proportional Fitting (IPF) to create baseline population

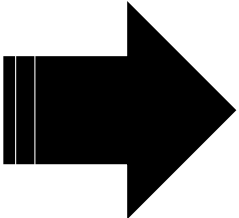
4. Assign PUMS records



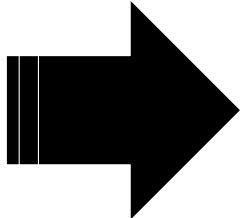
Optimal Results of New Pipeline

- Organized scripts and repository
- Use in projects that benefit from newer populations (opioid usage)
 - Ability to model population over time (student loans)
 - **Decrease in time required for these tasks:**

Creation of new years' populations
(2014-16, future years)

Weeks  1-2 days

Requests for specific variables
from existing populations

Days  Few hours

References

- Getty Images. Creative #939837488, <https://www.gettyimages.com/detail/photo/human-crowd-forming-a-usa-map-population-and-social-royalty-free-image/939837488>
- Getty Images. Creative #1058756860, <https://www.gettyimages.com/detail/photo/dark-earth-map-with-glowing-details-of-city-and-royalty-free-image/1058756860>
- Getty Images. Creative #610779718, <https://www.gettyimages.com/detail/illustration/vector-illustration-of-blue-and-yellow-royalty-free-illustration/610779718>
- “RTI U.S. Synthetic Household Population™.” *RTI*, RTI International, 2019, www.rti.org/impact/rti-us-synthetic-household-population%E2%84%A2.