

The Critical Role of Use Cases in Creating 21st Census *Curated Data Enterprise (CDE)* Infrastructure and Capabilities

Edward Wu & Stephanie Shipp

Vicki Lancaster, Joseph Salvo, and Cesar Montalvo

University of Virginia, Biocomplexity Institute

A4: Statistical Products First – A Paradigm Shift for Federal Statistics

FCSM 2023

October 24, 2023 @ 9 am

Supporting Census Bureau for 2030 and beyond

Extending Data Commons to create Census Curated Data Enterprise

- Census Bureau collaboration
- Alignment with Census Bureau's
 - IT Enterprise modernization and transformation
 - *Statistical Products First* approach



“We are moving towards a **single enterprise, data-centric** operation that enables us to **funnel data from many sources** in a single data lake using common collection and ingestion platforms ...

This is the **essence of a curated data approach — assemble, assess and fill in the gaps to create quality statistical data.**”

Rob Santos, US Census Bureau Director

Use Cases Guided by the Curated Data Enterprise



What are some purposes and uses?

Disseminating products to meet data user needs



Equity and Inclusion



Counting Children



Family Units



Rural Communities



Local Businesses



Tribal Areas



Vulnerable Populations



Elderly Residential Care



The Unhoused



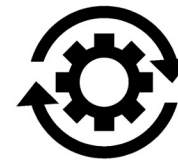
Migration



Climate Change



Gig Economy



Workforce Automation



Differential Undercount



Broadband Deployment



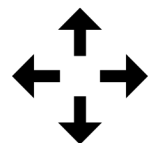
Emergency Preparedness



Economic Mobility



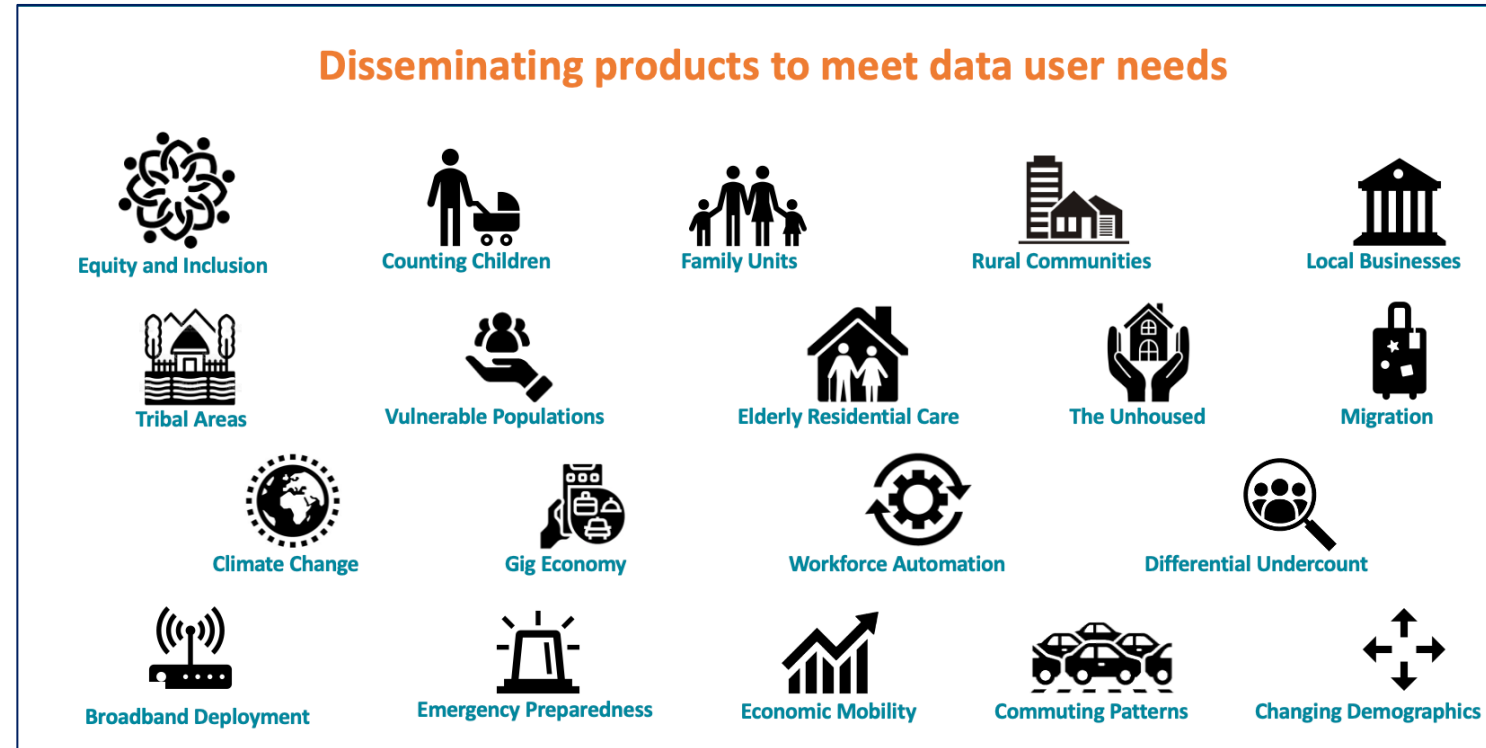
Commuting Patterns



Changing Demographics

Purposes and Uses

- **Research questions**
 - stakeholder-driven
 - grounded in practical applications
- Address demand for **timely and geographically granular data**
- **Integrating multiple data sources** can allow for more comprehensive statistical products



CDE Framework – Guiding Principles & Research Steps

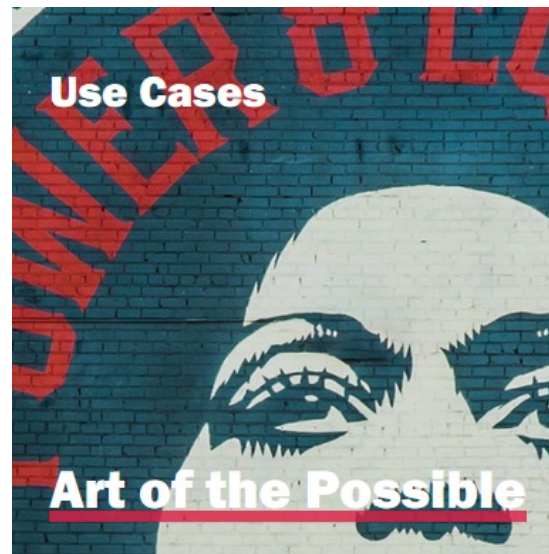
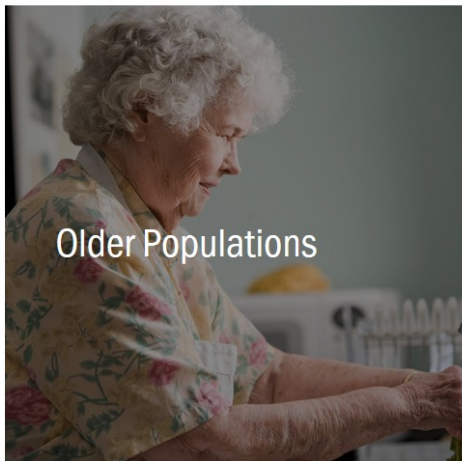
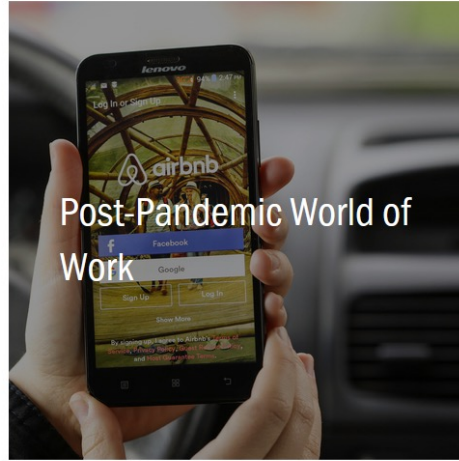
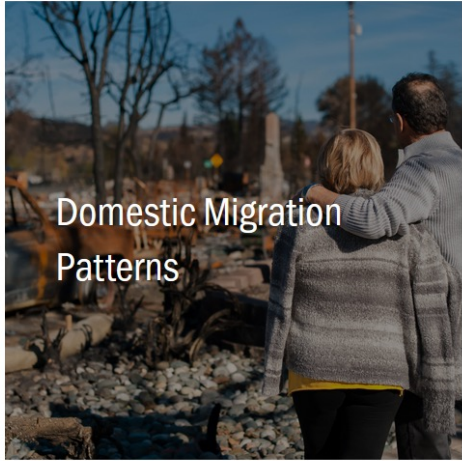
The Curated Data Enterprise approach goes beyond statistical products derived from single surveys, employing data re-use and integration to allow for timely and granular statistical products that help to answer pressing societal needs



Curated Data Enterprise Framework

Use Cases are
CDE Building Blocks
guided by the
CDE Framework.
1. Art of the Possible
2. Demonstration

Art of the Possible Use Cases – What are they?



Art of the Possible Use Cases

- Explore topics and related research questions that address pressing social needs identified by stakeholders
- **Characteristics**
 - Explore the current state of the topic
 - Discover data sources
 - Identify gaps and issues associated with existing data
 - Propose the feasibility of a **Demonstration Use Case**
 - Identifies the capabilities needed to develop and deploy the CDE

Art of the Possible Use Cases

- During **Listening Sessions**, we socialized the CDE concept with researchers and data users
- This feedback led to an initial set of *Art of the Possible* Use Cases that identified purposes and uses
 - Nursing Homes
 - Gig Economy
 - Migration due to climate events
 - Housing Affordability
 - Undercount of Young Children

Listening Sessions Stakeholder Groups

- Applied Demographers
- Advocates
- Economic Developers
- Economists and Business Leaders
- Public Policy Analysts
- Researchers

Topics

- Use Cases as CDE building blocks
- Curation of statistical product development
- Local Data Partnerships
- Third-party (private sector) Partnerships
- Use of ECIDS to address Undercount of Young Children

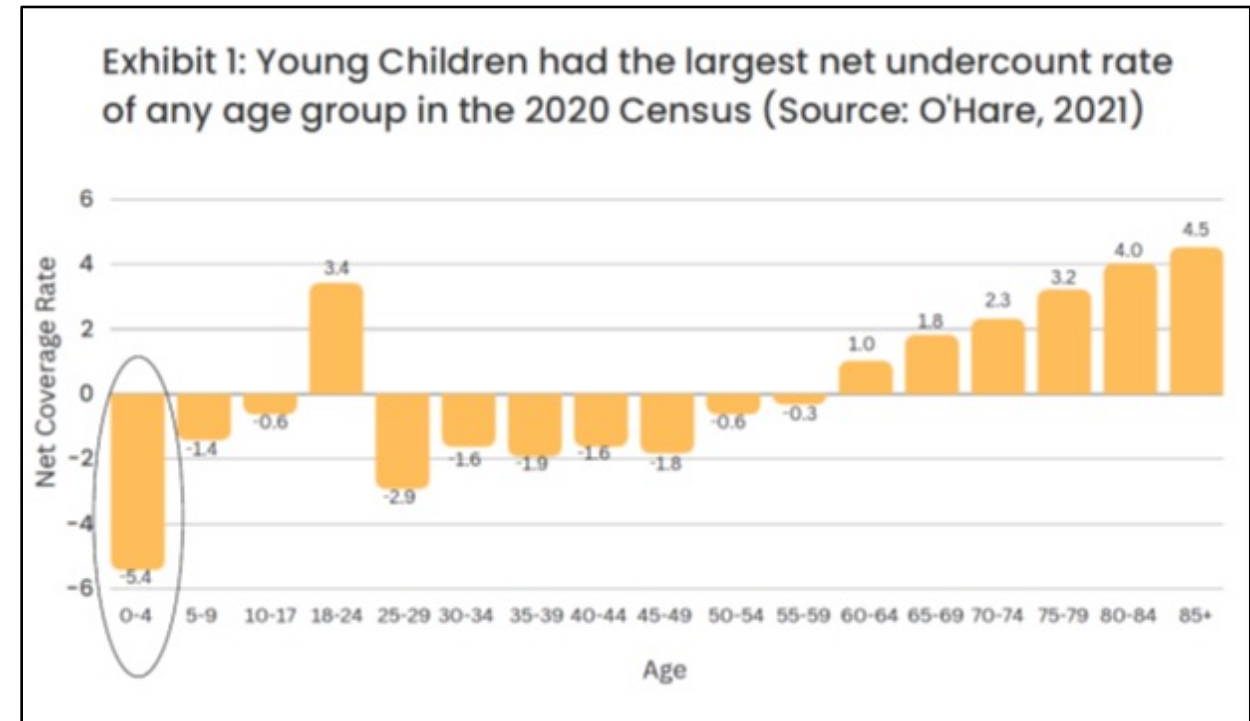
Addressing the Undercount of Young Children

Art of the Possible Use Case

Purposes and Uses

- **Baseline:** What data and methods can be used to improve the undercount of young children?
- What policy areas can be addressed with a better count of young children?
- How does the distribution of federal funds change if the counts of young children are more accurate?
- How successful is an early childhood vaccination program?
- What is the cost of implementing a pre-K program today and moving forward?
- How to best allocate grant resources to support childcare for mothers of young children?

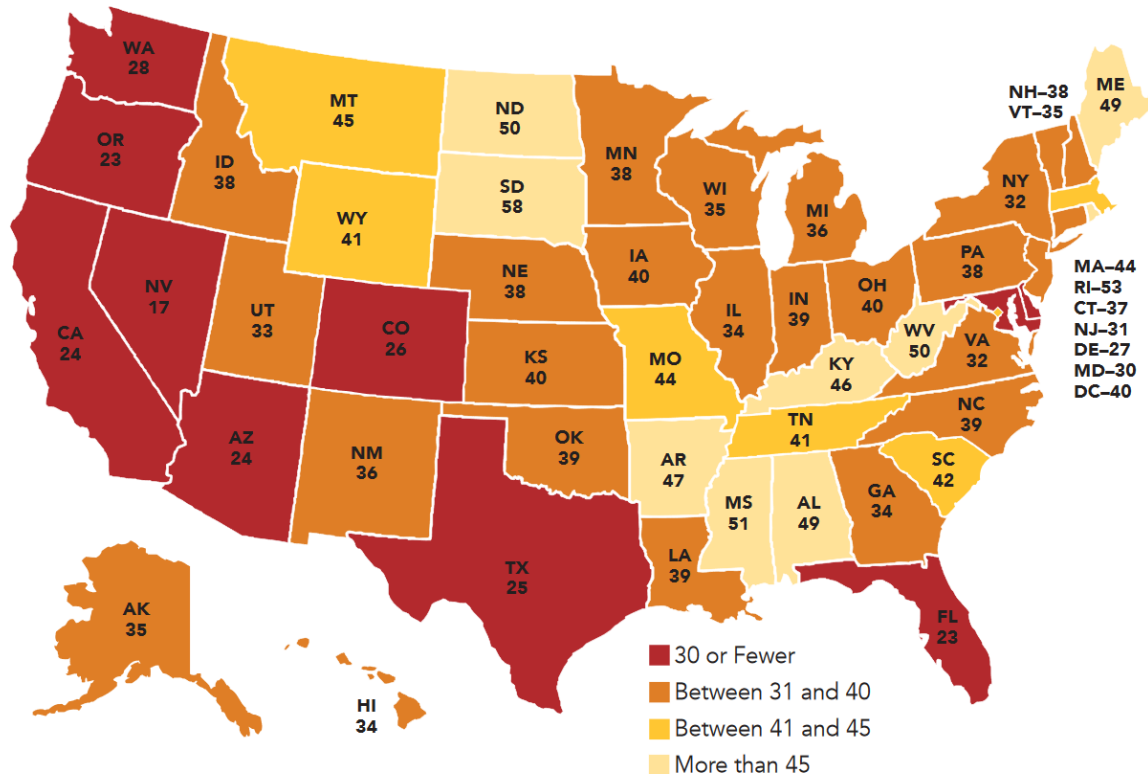
The **undercount of young children, ages 0-4**, is a persistent problem in the decennial census and the largest net undercount of any age group.



Rental Housing Affordability

Art of the Possible Use Case

National Low Income Housing Coalition Calculations for Rental Homes Affordable and Available per 100 Extremely Low-Income Renter Households



Note: Extremely low-income (ELI) renter households have incomes at or below the poverty level or 30% of the area median income. Source: NLIHC tabulations of 2021 1-Year ACS PUMS Data.

Issue

- There is a nationwide shortage of affordable housing
- Nearly half of renter households spend more than 30% of their income on housing

Purposes and uses - examples

- How would a change in **housing subsidies** affect the proportion of households burdened by housing costs?
- What are the effects of policies aimed to regulate **short-term rentals**?
- Which areas are most vulnerable to **increased energy bills** due to climate change? How will this affect the proportion of households burdened by housing costs?
- After accounting for income and wealth, are **eviction rates different across demographic groups**?

Domestic Migration Patterns

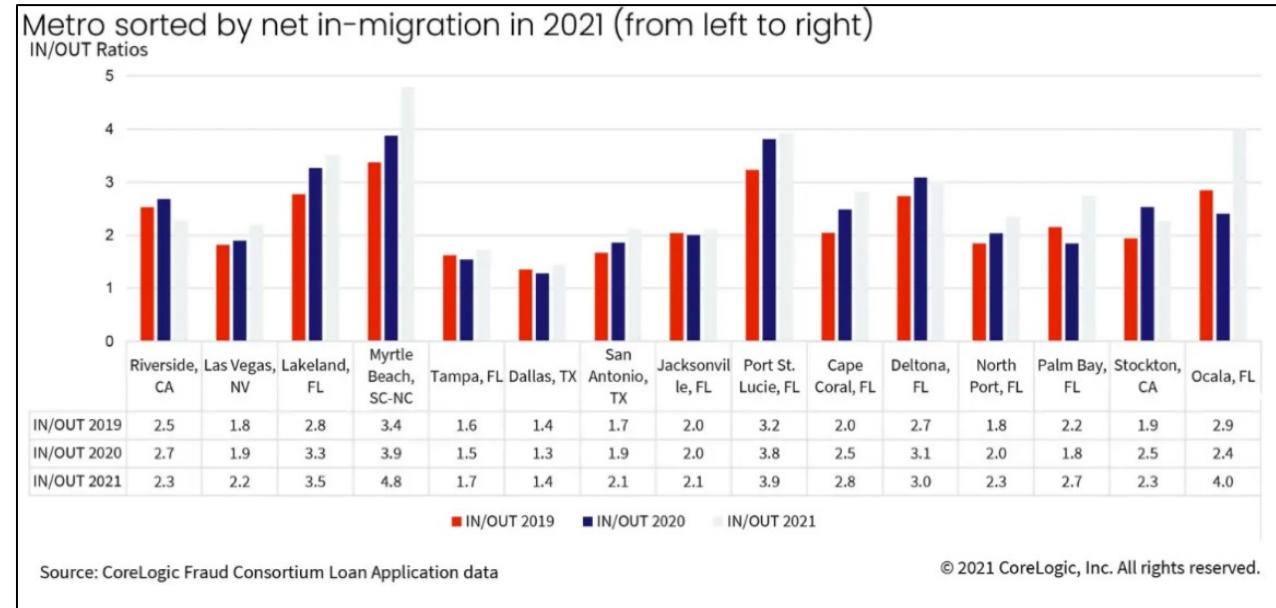
Art of the Possible Use Case

Migration is the most **volatile** component of population change

Purposes and Uses

- Baseline: What **methods and data** can be used to measure migration closer to real time?
- Can the **difference between temporary and permanent shifts** in population be identified?
- When households move due to climate events, **how equitable is access to services?**

Why, where, and how often do people move?

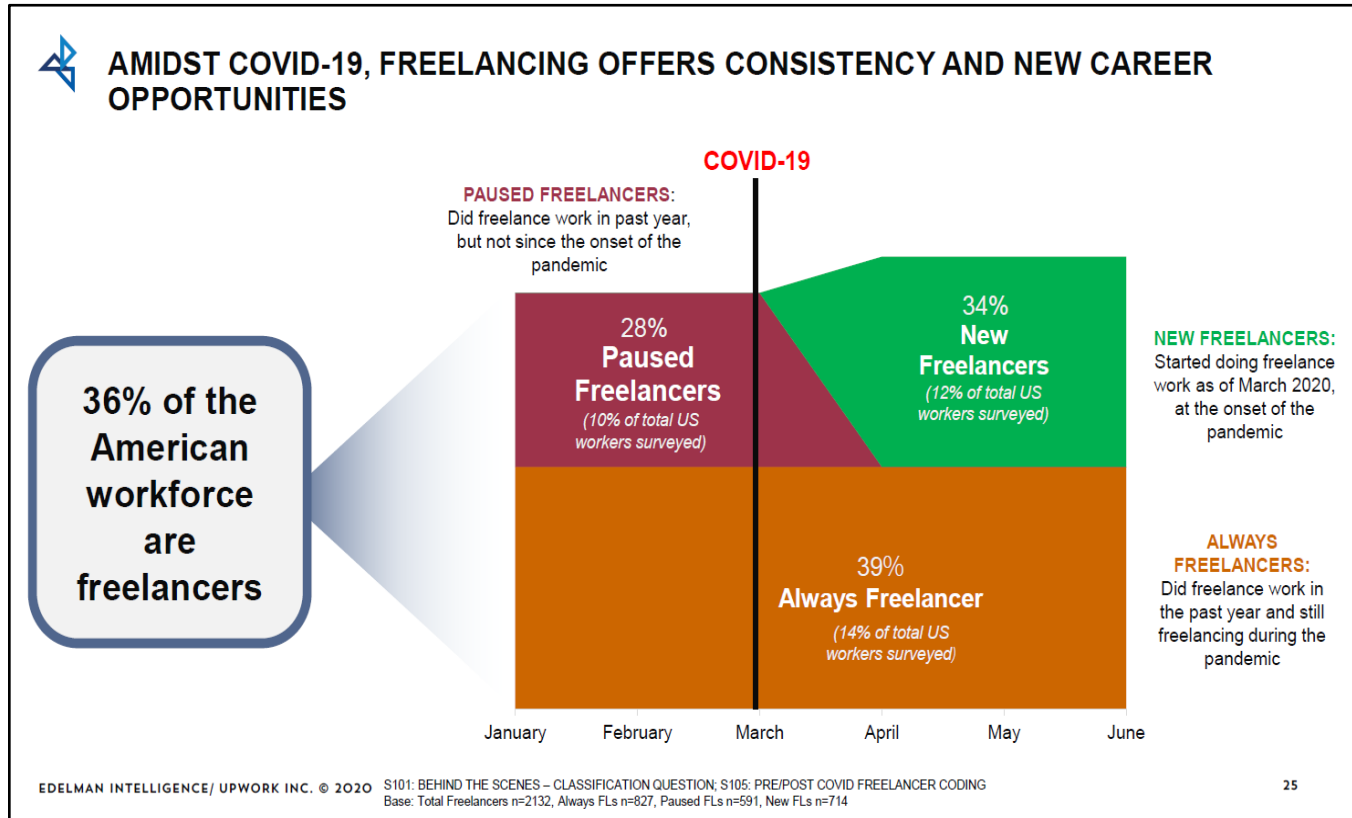


Share of House Hunters Looking to Relocate Is Near Pandemic Peak. *Percent of Redfin.com users searching for homes outside their home metro area.*

Measuring the Gig Economy

Art of the Possible Use Case

One estimate of Gig employment



Definitive estimates of gig employment arrangements vary widely depending on the definitions and sources used.

- Do not need to build from scratch
- Use an integrative approach
 - Can be informed by multiple studies, e.g., Abraham, et al, 2018
- Combine many data sources,
 - BLS Current Population Survey
 - BLS Contingent Worker Supplement
 - NORC General Social Survey,
 - Internal Revenue Service Administrative data
 - Census Bureau's non-employer data
 - Third-party data, e.g., Lyft, Uber, DoorDash

Skilled Nursing Homes' Emergency Preparedness for Flood Events

Demonstration Use Case

Purposes and Uses

Baseline question: Can we create a comprehensive Skilled Nursing Facilities data picture? (residents, nurses, owners, SNF building and safety characteristics)

Question 1: Can Skilled Nursing Facility workers make it to work during a flooding event?

Question 2: Are Skilled Nursing Facilities prepared for emergencies?

Question 3: Can the community provide support to Skilled Nursing Facilities during emergency events?

Capabilities developed

- **Data on residents, workers, and owners brought together**
- **Numerous administrative data** sources are discovered and used
- **Home location of SNF workers.**
 - Bias in the synthetic worker counts model and fitness-for-purpose evaluation.
 - SNF workers' transportation routes + risk of flooding by route.
 - Monte Carlo simulation to estimate the average probability of getting to work.
- **SNF Emergency Preparedness deficiencies**
 - SNF Life Safety Code (fire) deficiencies
 - Emergency preparedness and Fire Life Safety Code index evaluation and validation as measures of vulnerability
- **Community Resilience Indicators**
 - Number of shelter facilities and emergency service providers per 10,000 people.
 - Health professional shortage areas.

Household Living Budget at Census Tract Level

Demonstration Use Case

Purposes and Uses

Construct a household living budget to estimate the prevalence of economically vulnerable households at the census tract level.

1. Can we **construct defensible household living budget (HLB)** estimates for all household combinations at the census tract level?
2. Can we **create a synthetic population** using an iterative proportional fitting method with Public Use Microdata Sample (PUMS) household data and the American Community Survey (ACS) marginal distributions by household income and household size categories at the census tract level?
3. Can we **model economic vulnerability as a function of HLB** and household income, size, and composition in the census tract where the household resides?
4. Can we **validate** our estimates?

Capabilities developed

Discovered and used many novel and traditional data sources

HLB by household composition and size x census tract

- Component calculations by a variety of data sources
- Synthetic population
- Iterative Proportional Fitting results
- Visualizations
- Literature on budget standards, statistical methods, and the curated data environment

Documentation

- HLB Use Case GitHub Repository
- Data Tables and Derived Variables
- Technical report
- An application to food insecurity

CDE Use Case Research Program

Curated Use Case characteristics (to be tested)

- Engage the **research community**
- Promote **methodological research** to develop and implement CDE capabilities
- Develop an **end-to-end curation process** for the CDE
- Make **curated documentation available and searchable publicly (transparency)**



Program is designed to support the rapid development of statistical products

1. Criteria for selecting, deploying, and evaluating Use Cases to identify CDE capabilities
2. An End-to-End Curation process
3. Selection and invitation of researchers to participate in the CDE Use Case program.
4. Creating a research agenda for the US Census Bureau
 - Synthesize CDE capabilities
 - Propose statistical products
 - Assess the E2E Curation Process