Census Bureau Innovations:

Data Collection, Rapid Response, Integration, and Transparency

Jason Fields
U.S. Census Bureau

Outline

- Census Household Panel

- National Experimental Wellbeing Statistics (NEWS)

- Using Administrative Data in Editing and Weighting

- The Census Bureau / National Telecommunications and Information Administration (NTIA) ACCESS BROADBAND Dashboard

- The Community Resilience Estimates (CRE)

Themes

- Blending survey and non-survey data

- Improving data quality

- Moving administrative records from demonstration to production

- Innovations to create new products

- Transparency

Census Household Panel

- Outgrowth of the challenge and utility of the Household Pulse Survey (HPS)

- High-quality panel being developed to support collecting high-frequency data
  - improve representativeness
  - significantly reduce burden on households (relative to the HPS)

- A pool of households carefully selected, recruited, and refreshed by the Census Bureau to reflect the diversity of our Nation's population.

- The initial goals for the size of the Panel is 15,000 panelists.
  - Panelists initially sampled from and linked to the Census Bureau's gold standard Master Address File.
  - Representativeness will be enhanced by administrative data linked to the Master Address File
  - Refreshment will be facilitated by additionally linking respondents to the Demographic Frame by Protected identity keys (PIKs)

- This ensures the Panel is rooted in this rigorously developed and maintained frames and available for linkage to administrative records securely maintained and curated by the Census Bureau.

United States® Census Bureau

Census Household Panel

- Information collection online using Qualtrics as the data collection platform.

- Panelists will be able to respond online and by inbound CATI (computer-assisted telephone interviewing).

- Outbound CATI nonresponse follow-up will also be conducted.

- Panel recruitment will consist of mail contacts and telephone follow-up, with the possibility of personal visit.

- Participants will be enrolled via a screener and a baseline questionnaire.

- Panelists will be invited to complete monthly topical surveys.

- Incentives will be provided to respondents.

Census Household Panel

- This Panel will become integral to rapidly providing insight on national events that may impact social, economic, or demographic characteristics of the population.

- The Panel will also help us research questions related to surveys.

- Conduct nationally representative field tests to test content changes in an efficient and reliable fashion in support of other surveys.

- Alternative methods for enhancing data with administrative and other external data sources and develop modeled data.

- Critical platform for developing adaptive design procedures that use auxiliary data sources. Adaptive design has proven to reduce costs, improve data quality, and maintain and improve representativeness in the data we collect and use.

National Experimental Wellbeing Statistics (NEWS)

- Rethink how we can produce income and resource statistics
  - What is the best possible estimate given all the data currently available at Census for a given income/resource statistic?
  - Expand the set of income and resource statistics we produce
- V1 released – February 14
  - Proof of concept
    - 1 year
    - Mirror income and poverty releases – money income (no taxes, credits, in-kind benefits)
    - Present methods and approach for feedback
- Paper and estimates available at
  - https://www.census.gov/data/experimental-data-products/national-experimental-wellbeing-statistics.html
- Transparent
  - Code and data will be made available in RDCs, as much as possible
  - Tons of detail on steps in paper
  - Plan to release code publicly (subject to DRB constraints)

National Experimental Wellbeing Statistics (NEWS) - Measurement Challenges

Survey data

- Unit Nonresponse Bias
  - Not answering the survey
  - Poverty biased **down** by 0.3-0.5 percentage points during the pandemic (Bee and Rothbaum, 2022)
- Item Nonresponse Bias
  - Not answering income questions (~45 percent of income in the CPS ASEC is imputed!)
  - Poverty biased **down** by 0.5-1 percentage points (Bollinger et al., 2019; Hokayem et al., 2022)
- Mis- and underreporting
  - Not answering accurately
  - Poverty biased **up** by 2.5 percentage points for individuals 65+ (Bee and Mitchell, 2017)
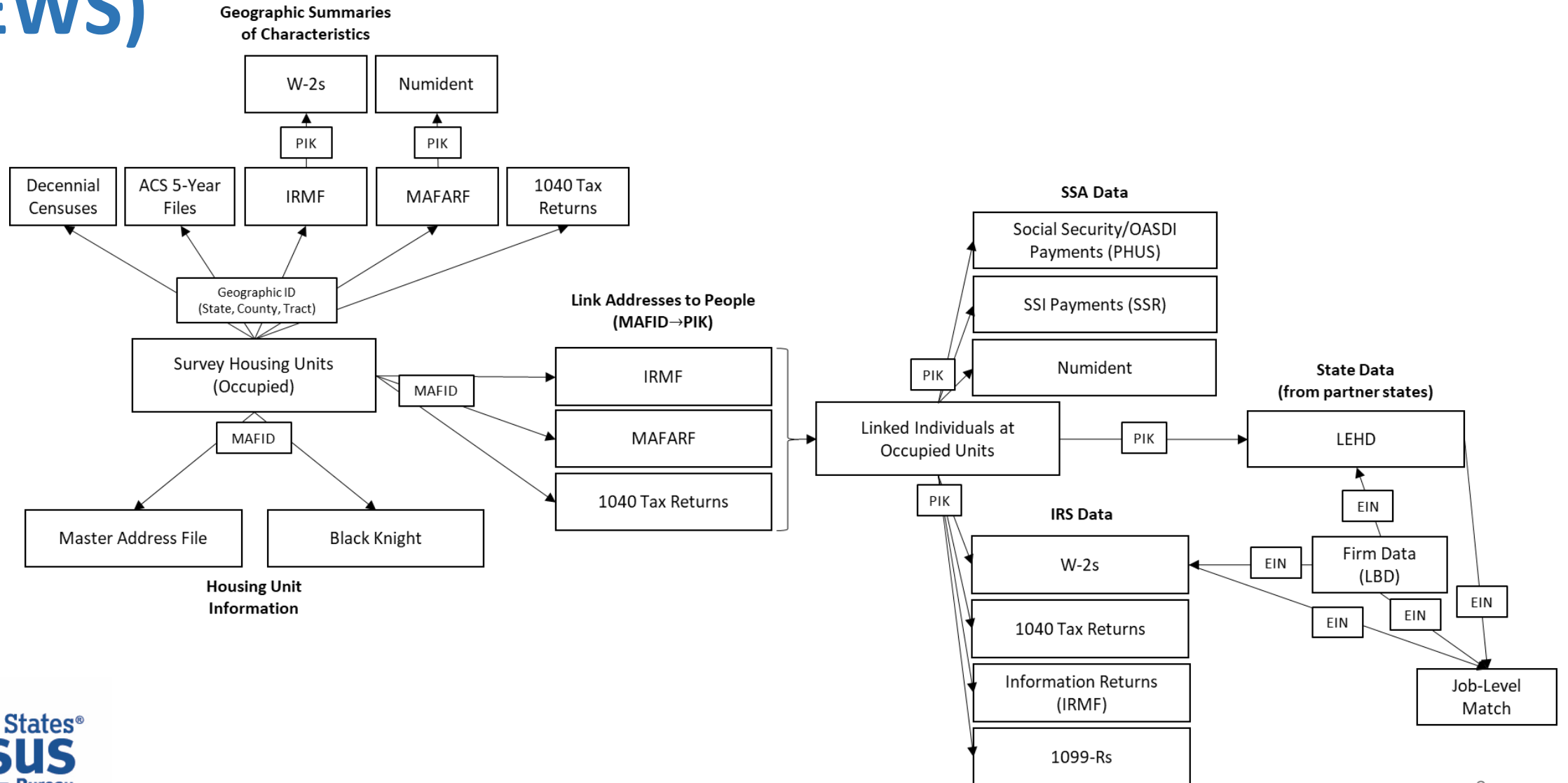
National Experimental Wellbeing Statistics (NEWS) - Measurement Challenges
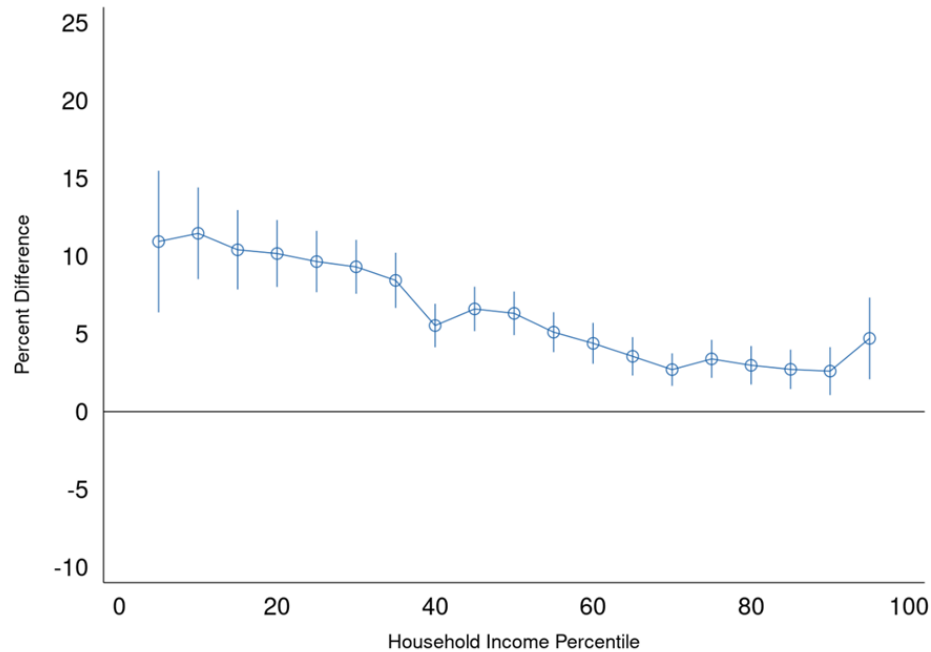
Administrative data

- Selection into administrative data
  - Not everyone has to file taxes or gets a W-2 or other information return
  - Larrimore, Mortenson, and Splinter (2020) estimate poverty from administrative data, but must impute the existence and poverty status of 4-6 million people
- Administrative data "nonresponse"
  - Some information not reported that should have been
  - Under-the-table jobs without a W-2, for example – 5% of adults in CPS ASEC report wage and salary earnings on the survey with no W-2
- Administrative mis- and underreporting
  - Not always 100% accurate
  - Unreported tips, underreported self-employment earnings (refer to IRS tax gap analyses)
- Conceptual misalignment
  - Administrative not always measuring what we want
  - W-2s historically do not have earnings used to pay for health insurance premiums – understate true earnings (Census also doesn't get this information when it's available)
- Incomplete data coverage
  - Data not available for individuals or places
- Selection into linkage
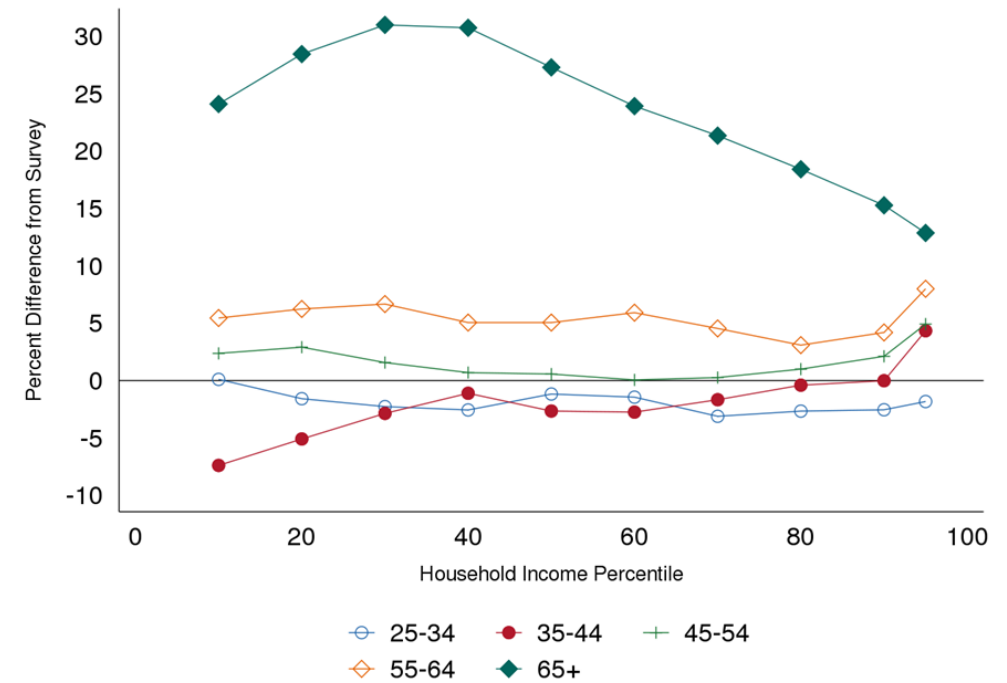 - Not all individuals can be linked across data sources (refer to Bond et al., 2014)

National Experimental Wellbeing Statistics (NEWS)

National Experimental Wellbeing Statistics (NEWS)

Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

Using Administrative Data in Editing

Hot-deck imputation
- Similar respondents are grouped together via stratifying variables
  - Race, sex, gender, marital status, and other demographic (categorical) variables
- Missing values imputed from similar respondents with non-missing values
- Disadvantages
  - Only being able control for limited characteristics; To create donor cells that were sufficiently large
  - Missing at random assumption may not hold

SRMI Model-based imputation
- Include large number of predictor variables
  - Continuous predictors allowed
  - Auxiliary data (e.g., IRS and SSA data) can be used
- Customizable regression models
  - Choose relevant regressors & stratifiers, regression type, skip patterns, etc.
- Preserve correlation across topics by estimating a joint distribution for imputation
- Meets Census Bureau disclosure standards
  - Predicted from regression models, not deterministic swapping of values from another person or another data source

Using Administrative Data in Editing

Model Types used in SIPP

Logistic Regression: binary variables (topic flags, most common)
- Social security benefits during the reference year
- SSI benefits during reference year

Linear regression: continuous variables
- Earnings or benefit amount
- Asset values

Multinomial regression: categorical variables (SIPP 2018+)
- Employment characteristics (business profit/loss)

Bayes Bootstrapping (donor-based): categorical/binary variables
- For adrecs/non-outputted variables (speeds up processing)

- *Note: SIPP demographic variables, which are used as regressors, are already hot-decked & edited prior to SRMI processing*

Also using SRMI in NSCH and developing models for Consumer Expenditure Survey, evaluations ongoing in CPS, ACS, and Decennial

Using Administrative Data in Editing

Educational Enrollment

Employment

Program Participation
- General Assistance
- Supplemental Nutrition Assistance Program (SNAP)
- Supplemental Security Income (SSI)
- Temporary Assistance for Needy Families (TANF)
- Women, Infants, and Children (WIC)

Health Insurance
- Private
- Medicaid
- Medicare
- Military
- Other

Biological Parent (fertility)

Disability
- Functional Limitations
- Difficulty Finding/Keeping Job

Liabilities
- Had credit card debt
- Had student loans
- Had medical debt
- Had other loans
- Amount of medical debt

Other Sources of Income
- Disability Payments
- Energy Assistance
- Lump Sum Payments
- Retirement/Retirement Payments
- Life Insurance
- School Breakfast and Lunch
- Social Security- Adults
- Social Security- Kids
- Survivor Payments
- Unemployment Compensation
- Veterans Affairs Benefits
- Worker's Compensation
- Payments to cover costs of dependent care

Retirement
- Received (contributed) income from (to) IRA
- IRA through main job
- Received income from 401k
- 401k from main job
- Made contributions to 401k
- Had a 401(k) through a previous employer
- Received income from pension
- Pension or retirement plan through main job
- Amount of income received (contributed) from (to) IRA
- Amount of income received (contributed) from (to) 401k

Assets
- Owned annuities or trusts
- Owned businesses as an investment only
- Owned certificates of deposit
- Owned checking accounts
- Owned gov't securities/savings bonds
- Owned IRA or KEOGH
- Owned annuities or trusts
- Owned municipal or corporate bonds
- Owned mutual funds
- Owned money market accounts
- Owned other financial investments
- Participated in a defined-benefit pension
- Owned other real estate
- Owned rental property
- Owned savings accounts
- Owned stocks
- Owned 401k, 403b, 503b, or TSP
- 401k balance
- IRA/Keough balance

Employment Characteristics
- Job(1-7) earnings
- Job(1-7) self employment
- Job (1-7) incorporation status
- Job (1-7) profit/loss

Using Administrative Data in Weighting

- Problems facing survey quality → Improve how survey weights are created to better address nonresponse bias

- What do survey weights do?
  - Simple Overview: Increase or decrease the "importance" of individual respondents to make the responding sample look more like the target population - (I.e. if older individuals are more likely to respond to a survey than younger individuals, then younger individuals would be given a higher weight value to make the sample more representative)

- Limitation: Current procedures only use demographic and geographic variables, but many key statistics from U.S. federal surveys are related to economic characteristics (e.g., median household income, unemployment rate)
  - Unclear how well demographic and geographic variables can correct for nonresponse bias with respect to economic characteristics

- However, through data sharing agreements, it is possible to link households (including nonrespondents) to administrative data from federal, state, and local agencies and third-party data sources

United States® Census Bureau

Using Administrative Data in Weighting

- Additional data sources beyond current survey controls provide more detailed information on how respondents and nonrespondents differ, which allows us to better correct for nonresponse bias

- Eggleston, Jonathan, and Ashley Westra. 2020. "Incorporating Administrative Data in Survey Weights for the **Survey of Income and Program Participation**." SIPP Working Paper No. 292. U.S. Census Bureau.
  - https://www.census.gov/library/working-papers/2020/demo/SEHSD-WP2020-07.html

- Little and Vartivarian (2005) "Does Weighting for Nonresponse Increase the Variance of Survey Means?"
  - Adding variables to weighting model that are correlated with both response likelihood and survey outcomes can reduce both bias and standard errors of estimates
  - Adding administrative data to weighting models could also help mitigate the effects of a sample cut on the precision of estimates

- The higher correlation between these administrative data and survey outcomes, the better these administrative data will be at improving data quality.
  - Best topics areas are ones where administrative data measure the same or a similar concept (E.g., Income, Poverty, Program Participation)
  - Medium benefit: Areas where there is high correlation, even if topic is not covered by administrative data (E.g., Census has no comprehensive administrative data on educational attainment, but income highly correlated with education)
  - Lower benefit: Topics that probably have low correlation with our administrative data. (E.g., Variables in CPS Volunteering and Civic Life Supplement)

NTIA/Census ACCESS BROADBAND Dashboard

Purpose:

In response to the ACCESS BROADBAND Act of 2021, the U.S. Census Bureau worked with NTIA to design a dashboard that will help policymakers and the public to assess how changes in broadband availability and adoption could influence local economies.

The ACCESS BROADBAND Dashboard consolidates multiple data sources in a mapping tool to enable researchers to assess changes in economic conditions based on changes in broadband availability and adoption, including those changes resulting from federal investments.

What is available in the dashboard?

The interactive dashboard includes a series of maps displaying statistics on broadband access, as well as select social and economic characteristics.

Map data is available for the following geographies:

- States (including the District of Columbia)
- Counties
- Census tract-level (limited)
- Puerto Rico (limited)

The dashboard includes 21 maps showing broadband access measures, as well as economic characteristics research suggests could be influenced by increased broadband availability and adoption
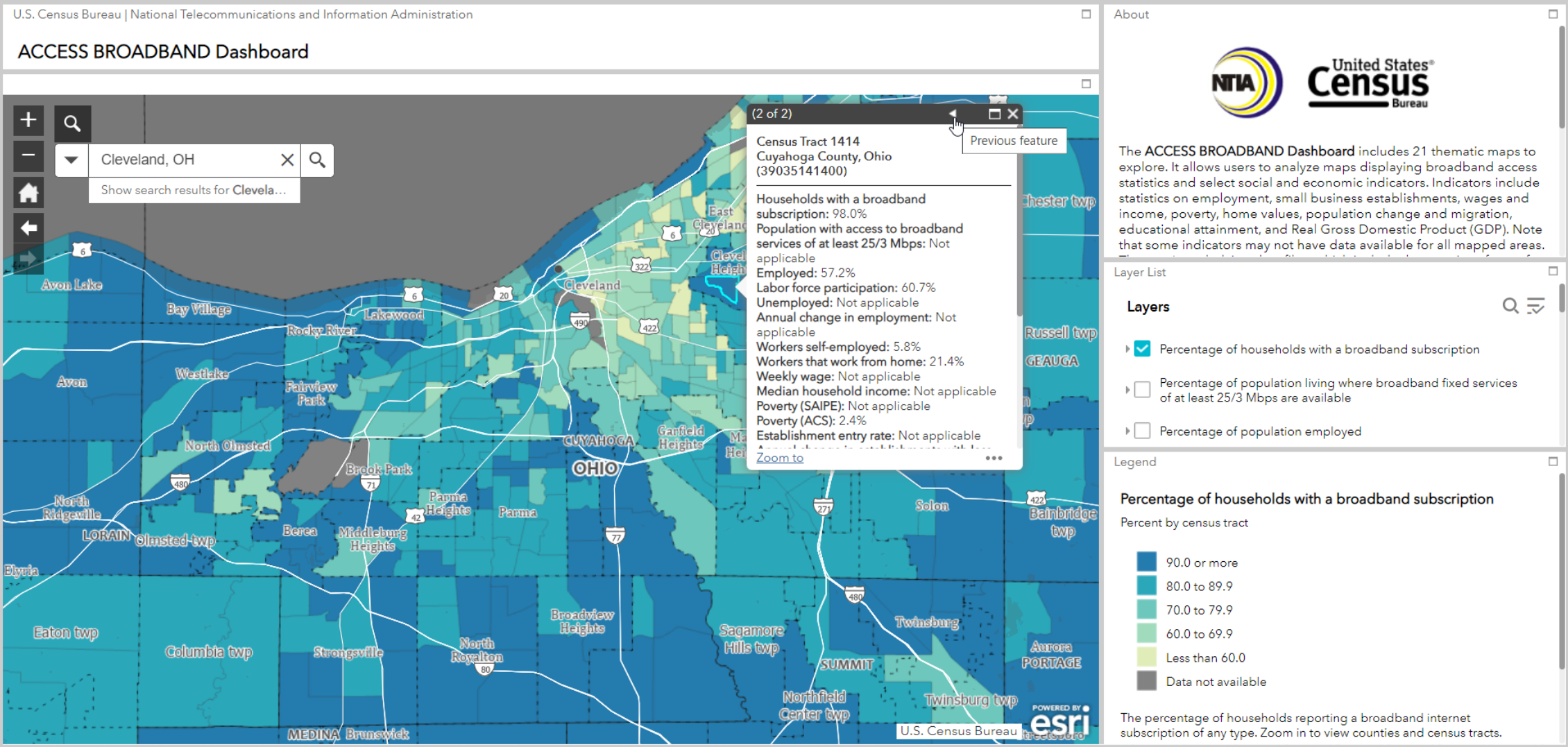
Layer List

Layers

- ☑ Percentage of households with a broadband subscription
- ☐ Percentage of population living where broadband fixed services of at least 25/3 Mbps are available
- ☐ Percentage of population employed
- ☐ Labor force participation rate
- ☐ Unemployment rate
- ☐ Annual change in employment
- ☐ Percentage of workers self-employed
- ☐ Percentage of workers that work from home
- ☐ Average weekly wage (dollars)
- ☐ Median household income (dollars)
- ☐ Poverty rate (SAIPE)
- ☐ Poverty rate (ACS)
- ☐ Establishment entry rate
- ☐ Annual change in business establishments with less than 20 employees
- ☐ Annual change in business establishments with less than 500 employees
- ☐ Annual change in Real Gross Domestic Product (GDP)
- ☐ Median home value (dollars)
- ☐ Annual change in population
- ☐ Net migration rate
- ☐ Percentage of population with a bachelor's degree or higher

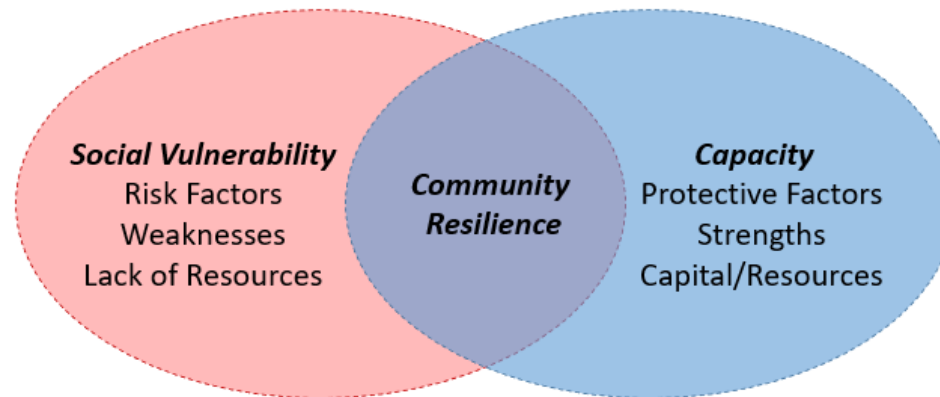https://www.census.gov/programs-surveys/community-resilience-estimates/partnerships/ntia/broadband-act.html

NTIA/Census ACCESS BROADBAND Dashboard

https://www.census.gov/programs-surveys/community-resilience-estimates/partnerships/ntia/broadband-act.html

Community Resilience & Social Vulnerability

▶ ***Social vulnerability*** is the risk of hazards to the physical and socially built environment, while ***community resilience*** is the capacity of individuals and households to absorb the stresses from a disaster.[2]

   ■ To eliminate the need to classify characteristics of an area as contributing to either vulnerability or resilience, resilience and vulnerability are viewed to represent two sides of the same resilience coin.[3]

▶ Social vulnerability mapping strengthens community resilience and reduces inequities.[1]

   ■ By helping communities better anticipate, respond, resist, and recover from disasters.

Social Vulnerability
Risk Factors
Weaknesses
Lack of Resources

Community Resilience

Capacity
Protective Factors
Strengths
Capital/Resources

1 Van Zandt, Peacock, Henry, Grover, Highfield and Brody 2012; 2 Masterson, Jamie Hicks, Walter Gillis Peacock, Shannon Zandt, Himanshu Grover, Lori Field Schwarz, and John Cooper 2014; 3 Summers, Smith, Harwell, and Buck 2017

United States® **Census** Bureau

https://www.census.gov/programs-surveys/community-resilience-estimates.html

Community Resilience Estimates (CRE)

▶ KEY BACKGROUND

- ■ Released in June 2020 in response to COVID-19 pandemic

- ■ Uses **restricted microdata** unavailable outside of the Census Bureau

- ■ Estimates are more **timely, accurate, reliable and granular** than anything else available

  -Uses 1-year ACS instead of 5-year ACS

  -Uses small area modeling to reduce sampling error by 30%

  -Complete coverage at the tract level

- ■ Methodology is **flexible** to other research interests beyond social vulnerability (e.g., broadband access, federal assistance eligibility)

https://www.census.gov/programs-surveys/community-resilience-estimates.html

Community Resilience Estimates (CRE)

▶ In addition to data files, CRE has interactive mapping tools

- Highlight areas with higher rates

- Interactive maps allow for dashboards with additional indicators

https://www.census.gov/programs-surveys/community-resilience-estimates.html
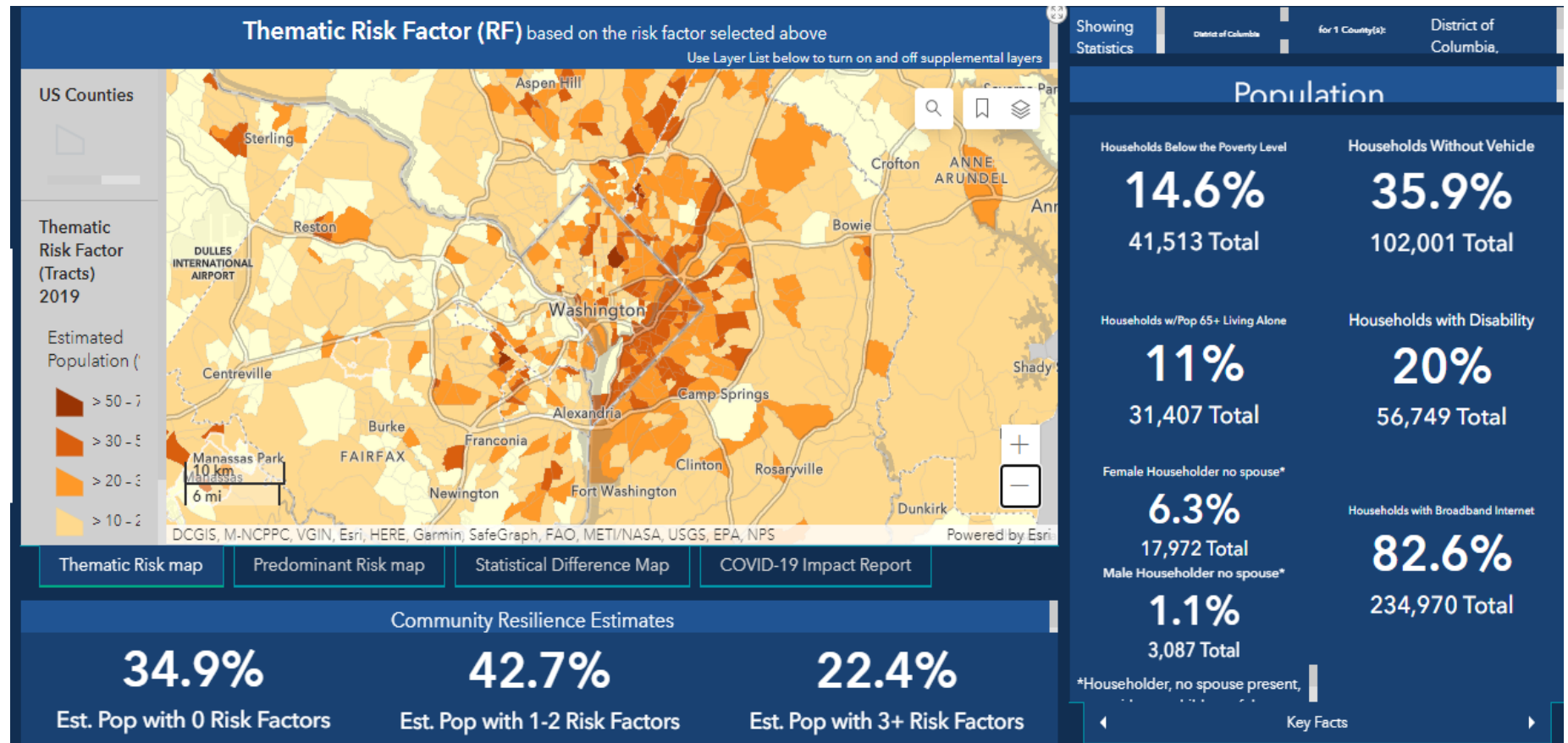
Thank you!

Please send any questions or comments, and
we'll get them to the right staff

Jason.M.Fields@Census.Gov