Balancing Respondent Confidentiality and Data User Needs

Aaron E. Cobet

Consumer Expenditure Surveys

Microdata Users Workshop

July 17, 2019

What is the Issue?

- Conflicting goals
  - Maximize data access
  - Protect respondents identity

Why is Confidentiality Important?

- Ensure trust of respondents for their cooperation
- It's the law

What is Title 13?

- U.S. Code: Title 13 allows the Census Bureau to take a survey and provides directives for its administration and enforcement.
- People who took the oath who wrongfully disclose information protected under Title 13 are subject to a fine of up to $250,000 or up to 5 years in prison or both.
- Census and CE staff need Title 13 clearance.

Title 13 Training

- CE staff gain access to internal data *after* completing 2 steps:
  1. Pass a background check by Census
  2. Take the Title 13 training
- CE staff are required to annually retake Title 13 training and pass a knowledge check to maintain Special Sworn Status

Who Determines Disclosure Threats?

Disclosure Review Board
of the Census Bureau

How Could Microdata Reveal Respondents' Identity?

Unique data points

- Names
- Addresses
- Extreme income

How to Protect Respondents' Confidentiality?

Conceal revealing information

- Census removes *direct* identifiers, e.g. names
- BLS suppresses *indirect* identifiers, e.g. high income

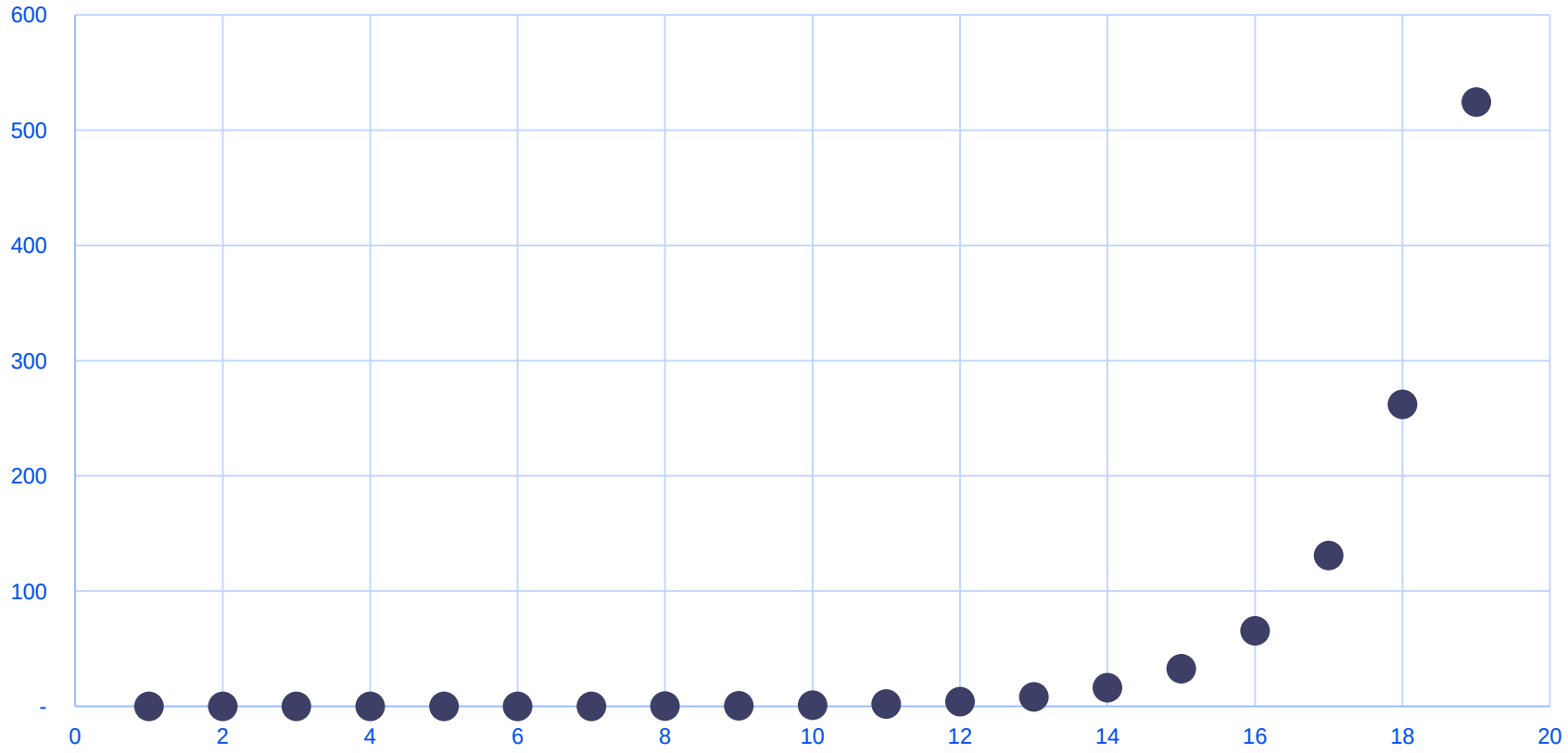
How to Conceal Indirect Identifiers?

- **Topcode**: Average numerical values above threshold
- **Recode**: Change item or CU characteristics
- **Suppress**: Delete numerical value or delete entire record

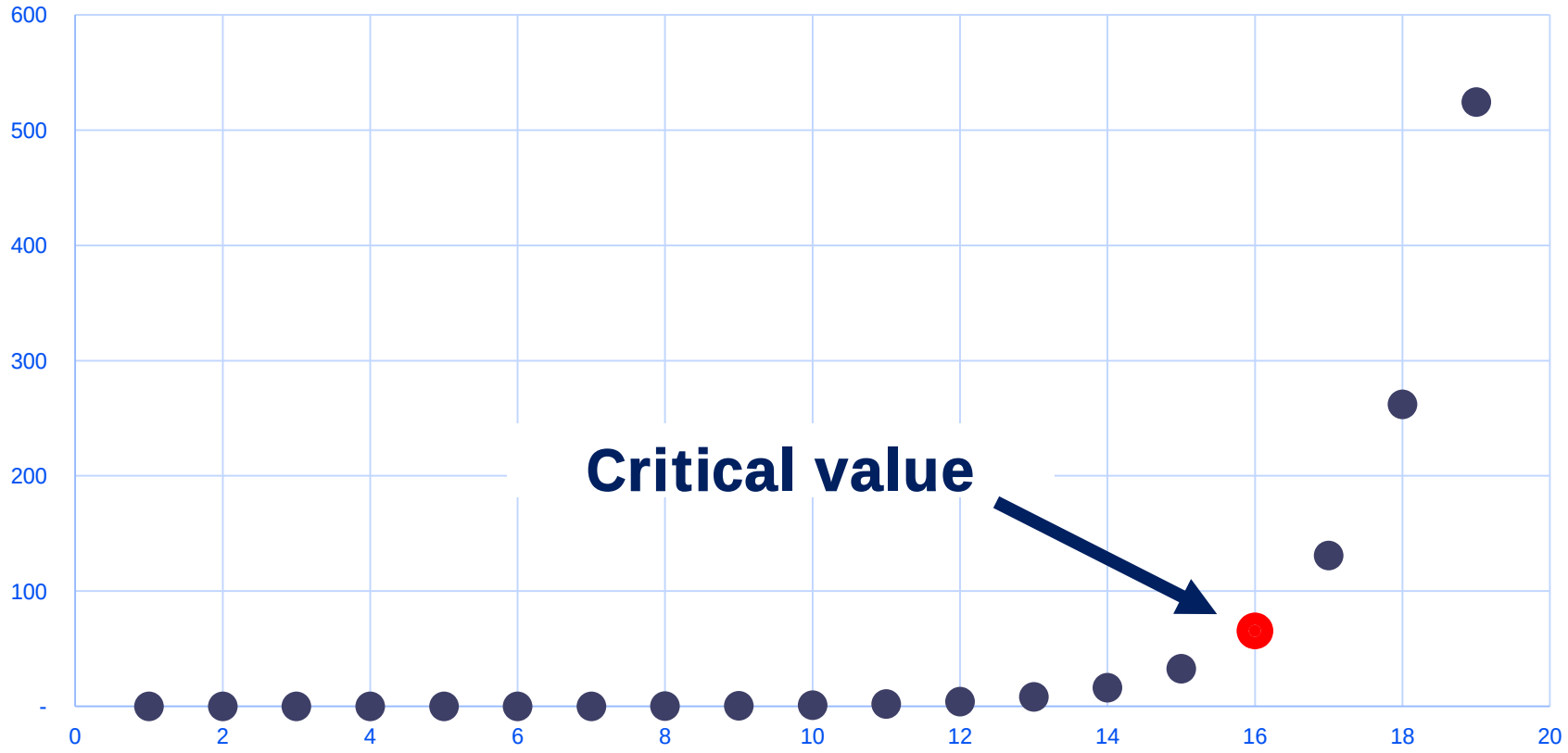
How do we Topcode?

- Determine critical value
- Average values exceeding critical value
- Replace exceeding values with top-coded values

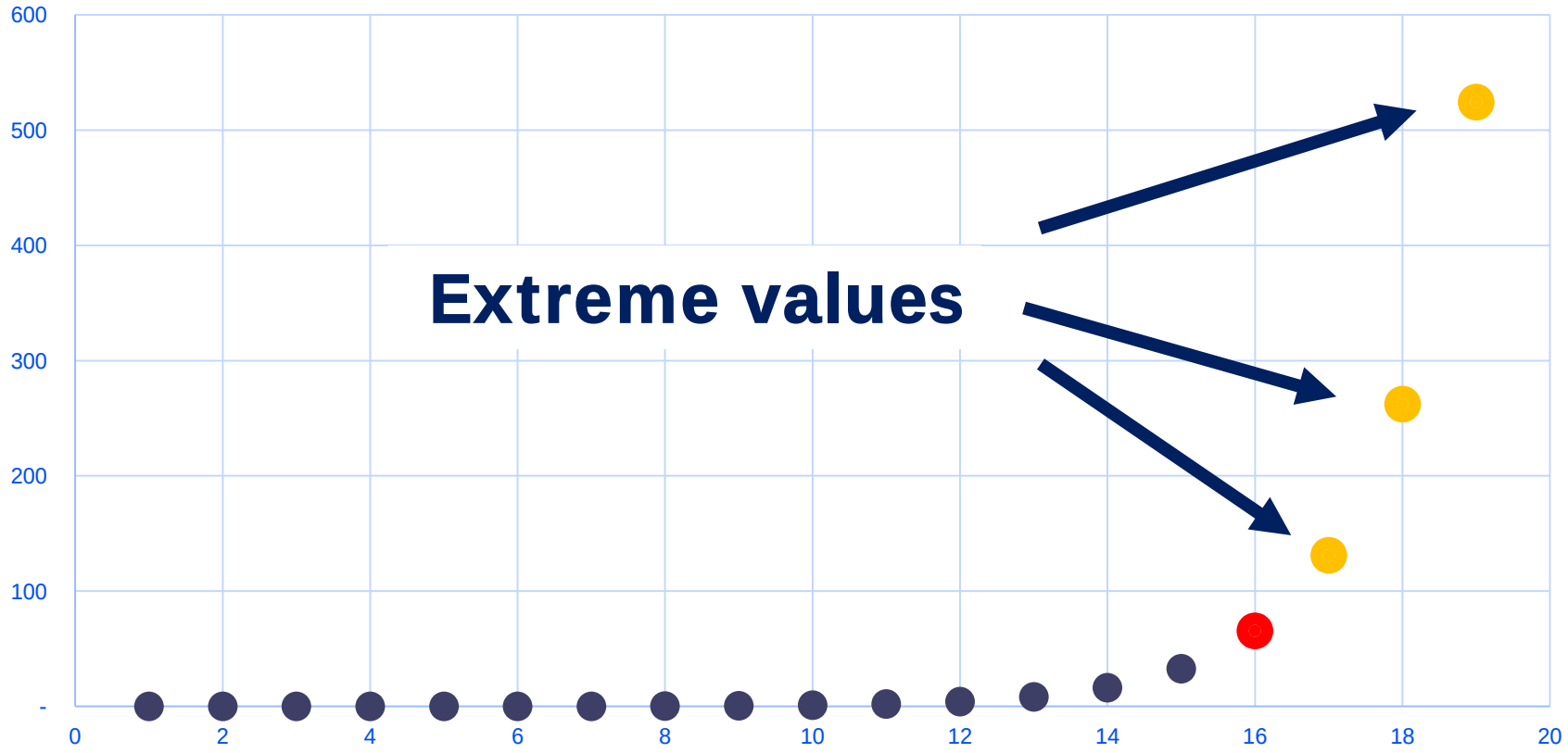
Topcoding Example



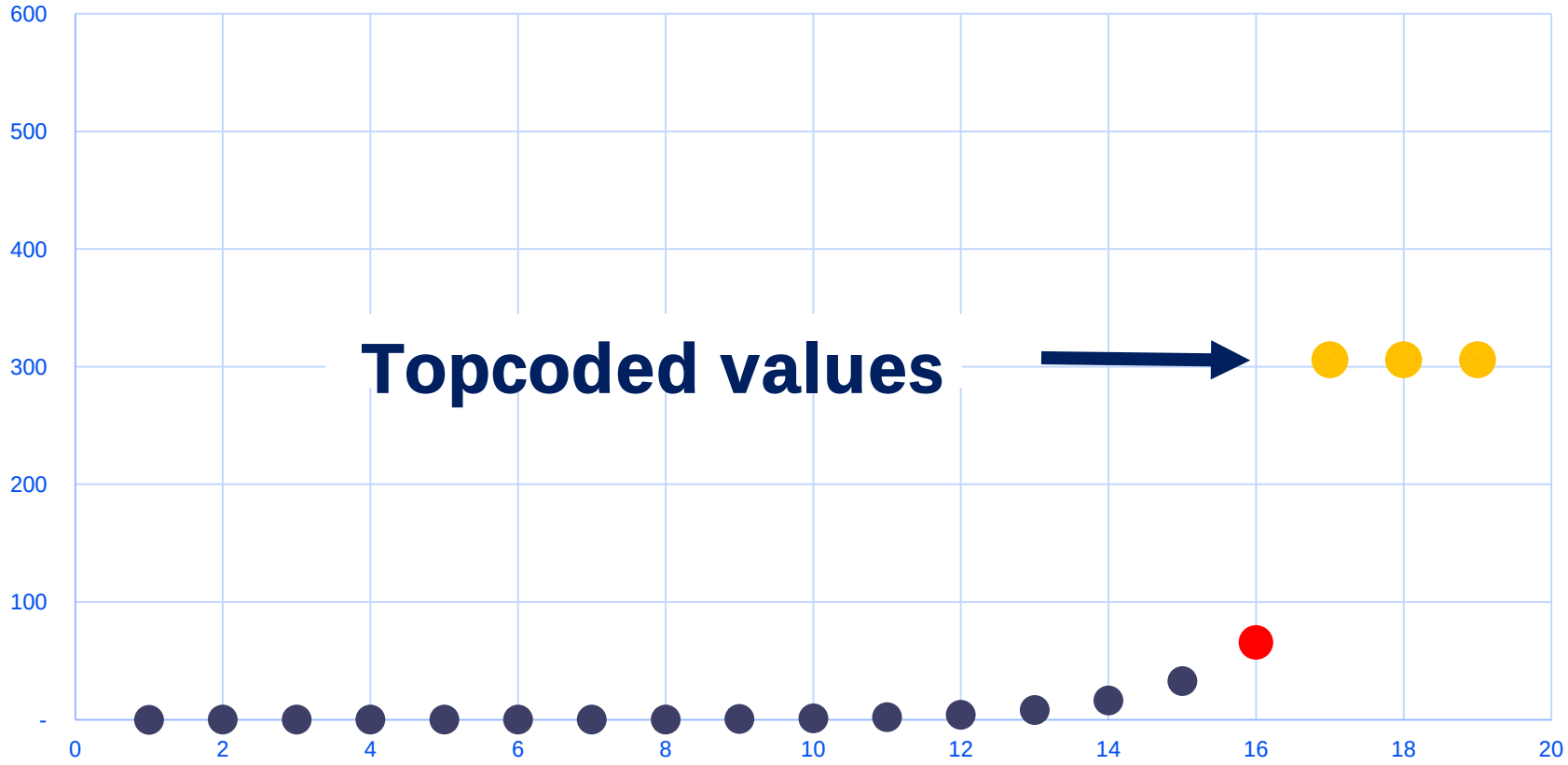
Topcoding Example



Topcoding Example



Topcoding Example



How to Determine Critical Values?

- Critical value is any value by a consumer unit above the specified percentiles:
  - Expenditures: 99.5 %
  - Income: 97.0 %

How do we Recode?

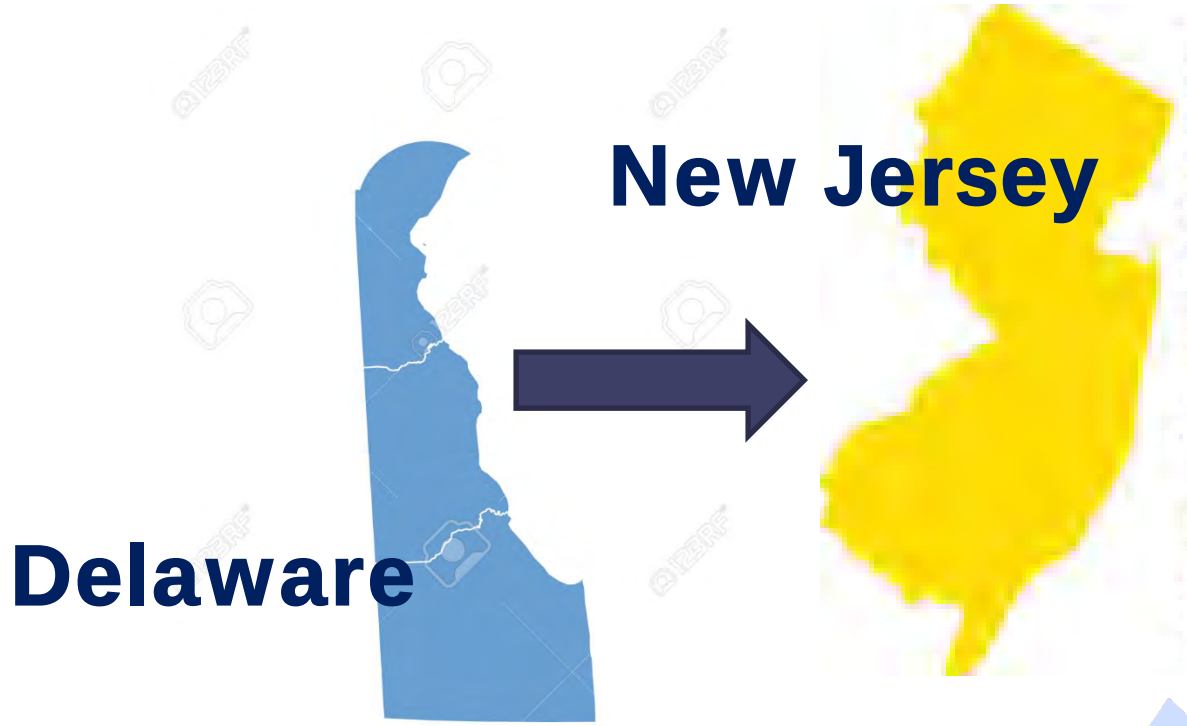
- Find revealing metadata
- Determine method:
  - Generalize information
  - Change information
- Replace original metadata with recoded metadata

Re-coding: Generalize Information

- From Toyota Corolla 1999
- To Toyota 1990s

Re-coding: Change information

- Change states to comparable states



How to Conceal Indirect Identifiers?

- **Top-coding**: Provide average of expenditures above a threshold
- **Re-coding**: Change item or CU characteristics
- **Suppression**: Delete numerical data or entire record

Suppression

- Erase aspect of the record
  - Example: State suppression
  - Example: Boat purchase
- Exclude entire record
  - Example: Airplane purchase

Reverse Engineering

$$5 = 3 + X$$

How to Prevent Reverse Engineering?

Prevent users to deduce protected information within files and across files

1. Find protected values
2. Protect them in all locations
3. Protect related values

Reverse Engineering: Within File

- Income = Wages + taxes
- 1000 = 800 + 200
- 1000 = 750 + 200
- 950 = 750 + 200

- Critical value: 700
- Topcode value: 750

Wages **exceeds** the critical value

Reverse Engineering: Within File

- Income = Wages + taxes
- 1000 = 800 + 200
- 1000 = 750 + 200
- 950 = 750 + 200

- Critical value: 700
- Topcoded value: 750

Wages **match** the topcoded value

Reverse Engineering: Within File

- Income = Wages + taxes
- 1000 = 800 + 200
- 1000 = 750 + 200
- 950 = 750 + 200

Wages and taxes **match** the income

- Critical value: 700
- Topcode value: 750

Reverse Engineering: Across Files

- **Income:** Topcoded income in FMLI
  - Topcode associated UCCs in ITBI
- **Expenditure:** Topcoded expenditures in EXPN and FMLI
  - Topcode associated UCCs in MTBI

How Do We Document?

Flag values

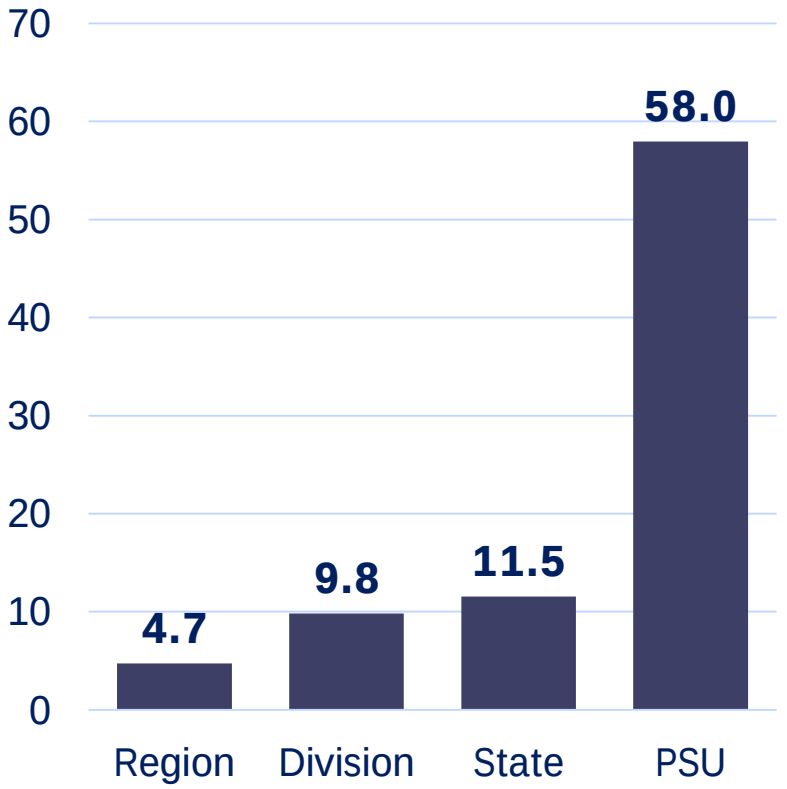
- **T**: Topcoded value
- **D**: Valid value, unadjusted

Impact of topcoding

- CE topcodes few observations
- Most affected data slices:
  - Geographic data non-self representing cities
  - Income for high earners

Impact of Suppression of Geographic variables, Percent

- Almost 60 % of suppressed PSUs
- Below 15 % of suppressed states, divisions, and regions



Source: FMLI and FMLD files for 2015.

Need More Data?

- Visiting researcher program
  - Access to pre-topcoded CE microdata
  - Requires application process
  - www.bls.gov/rda/home.htm

Additional Information

- Protection of respondent confidentiality (www.bls.gov/cex/pumd_disclosure.htm)
- PUMD Getting Started Guide (www.bls.gov/cex/pumd-getting-started-guide.htm)
- Title 13 (www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html)

Thank you!

Aaron Cobet

Senior Economist, Consumer Expenditure Surveys

(202)-691-5018

Cobet.Aaron@bls.gov