

Balancing Respondent Confidentiality and Data User Needs

Arcenis Rojas
Consumer Expenditures Survey
Microdata Users Workshop
July 13, 2016

What is the Issue?

■ Conflicting goals

- ▶ Maximize data access
- ▶ Protect respondents identity



Why is Confidentiality Important?

- Ensures future cooperation by respondents
- It's the law



What is Title 13?

- U.S. Code: Title 13 allows the government to take a census and provides directives for its administration and enforcement.
- Also says any person having taken the oath of office who wrongfully discloses information protected under Title 13 is subject to a fine of up to \$250,000 or up to 5 years in prison.

Title 13 Training

■ CE staff gain access to internal data *after* they complete 2 steps:

1. Pass a background check by Census (Special Sworn Status)
2. Take the Title 13 training

CE staff are required to annually retake Title 13 training and pass a knowledge check.

Who Determines Disclosure Threats?

- Disclosure Review Board (DRB)
by the Census Bureau



How Could Microdata Reveal Respondents' Identity?

- High expenditures
- High income
- Small PSUs

How to Protect Respondents' Confidentiality?

- BLS and Census Bureau conceal information that *could* reveal respondents identity.



How to Protect Respondents' Confidentiality?

Two stages:

- Census removes *direct* identifiers, i.e. addresses
- BLS suppresses *indirect* identifiers, i.e. high expenses



How to Protect Respondents' Confidentiality?

- **Topcoding:** Provide average of expenditures above threshold
- **Recoding:** Change metadata but provide numerical data
- **Suppression:** Delete numerical data or entire record

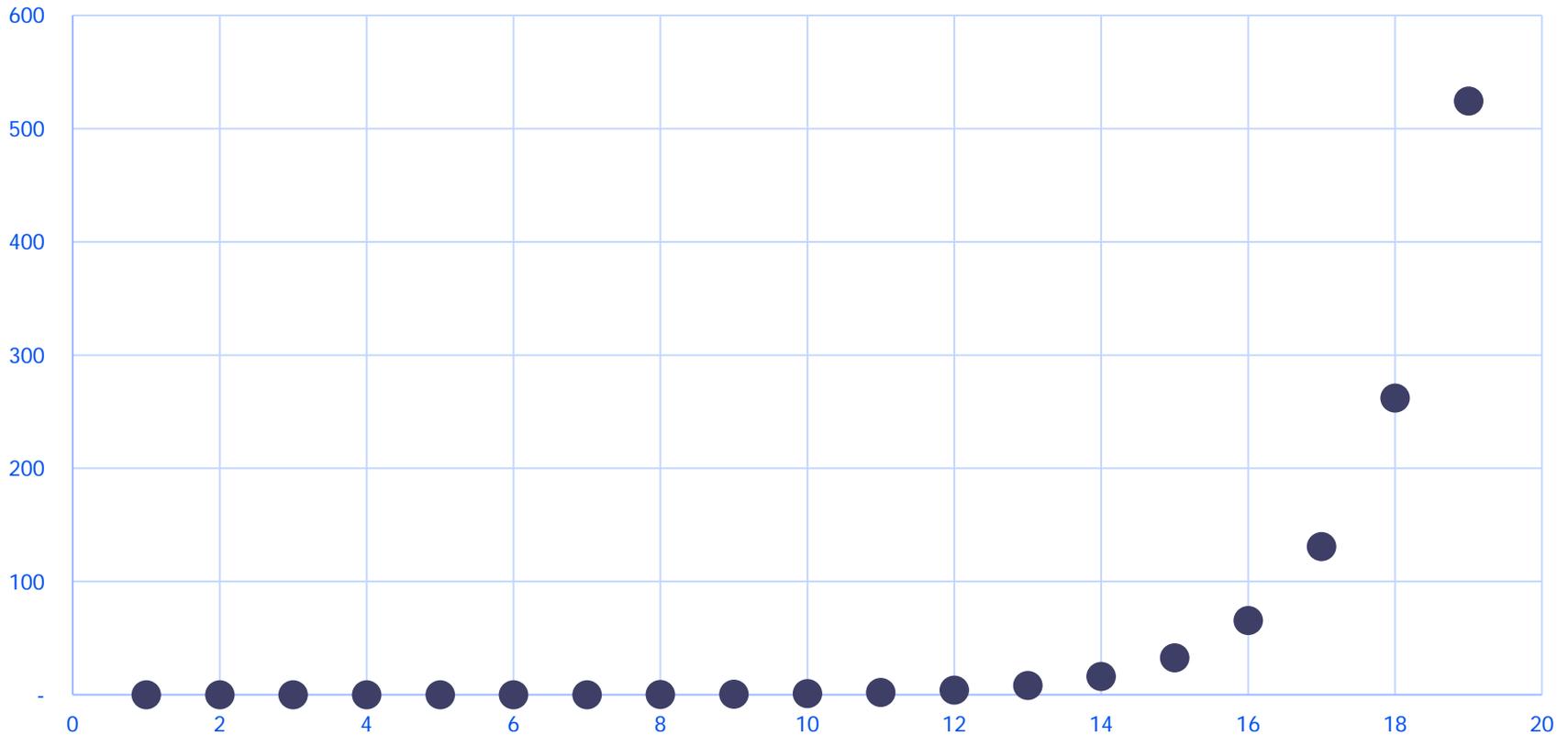
How to Protect Respondents' Confidentiality?

- **Top-coding:** Provide average of expenditures above a threshold
- **Re-coding:** Change metadata but provide numerical data
- **Suppression:** Delete numerical data or entire record

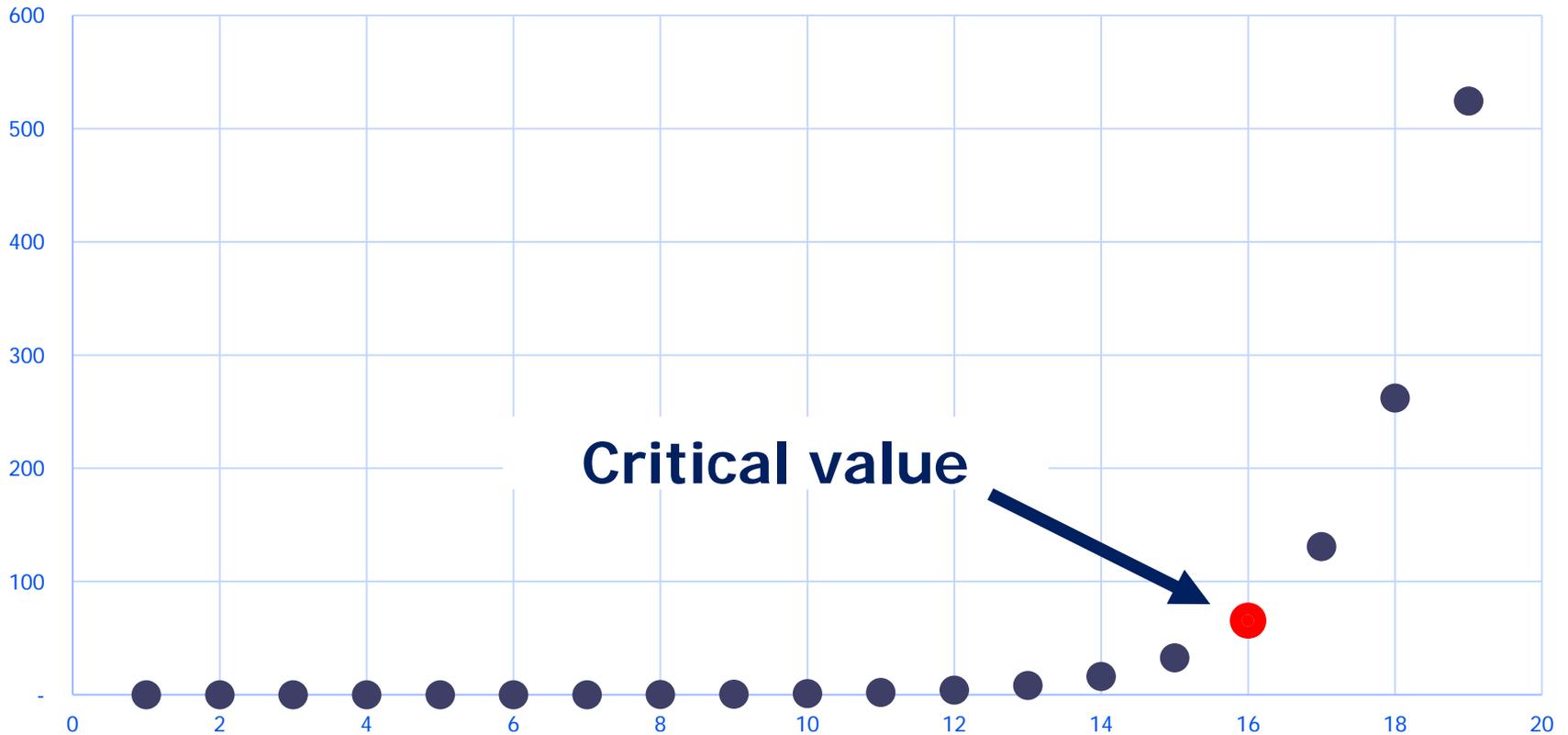
How Do We Topcode?

- Determine critical value
 - Find values exceeding critical value
 - Average values exceeding critical value
 - Replace values with top-coded values
-
- Bottom coding uses the same procedure but with negative values

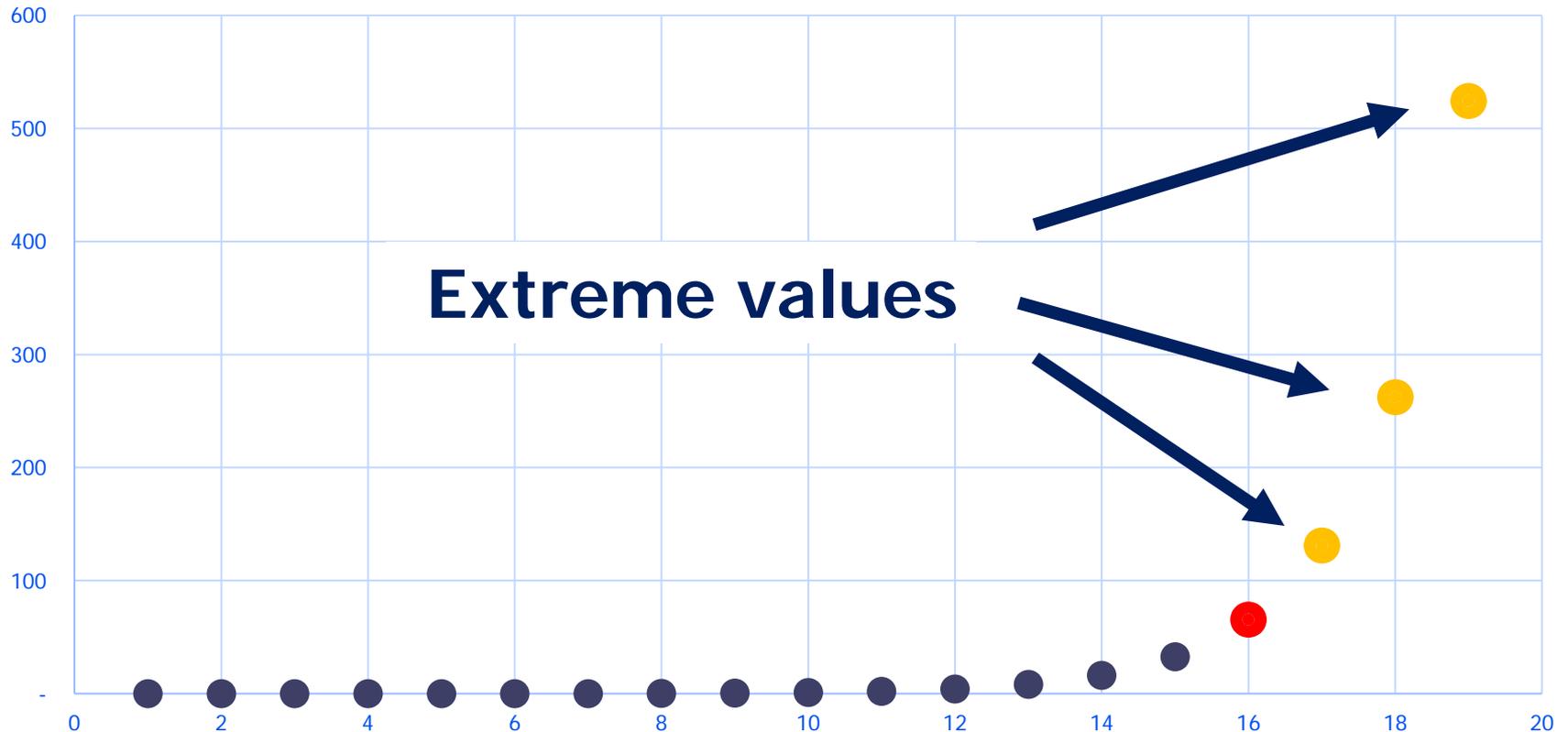
Topcoding Example



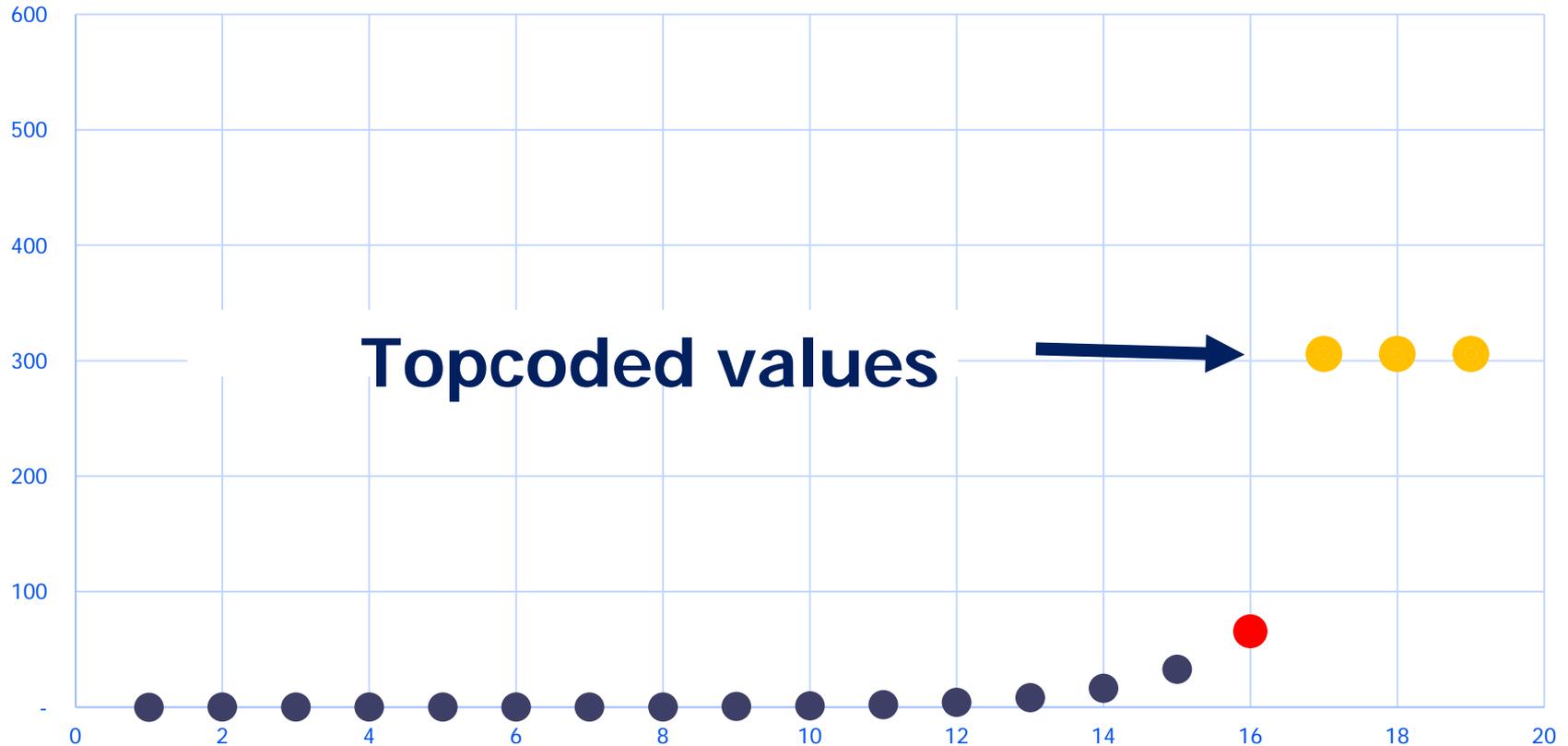
Topcoding Example



Topcoding Example



Topcoding Example



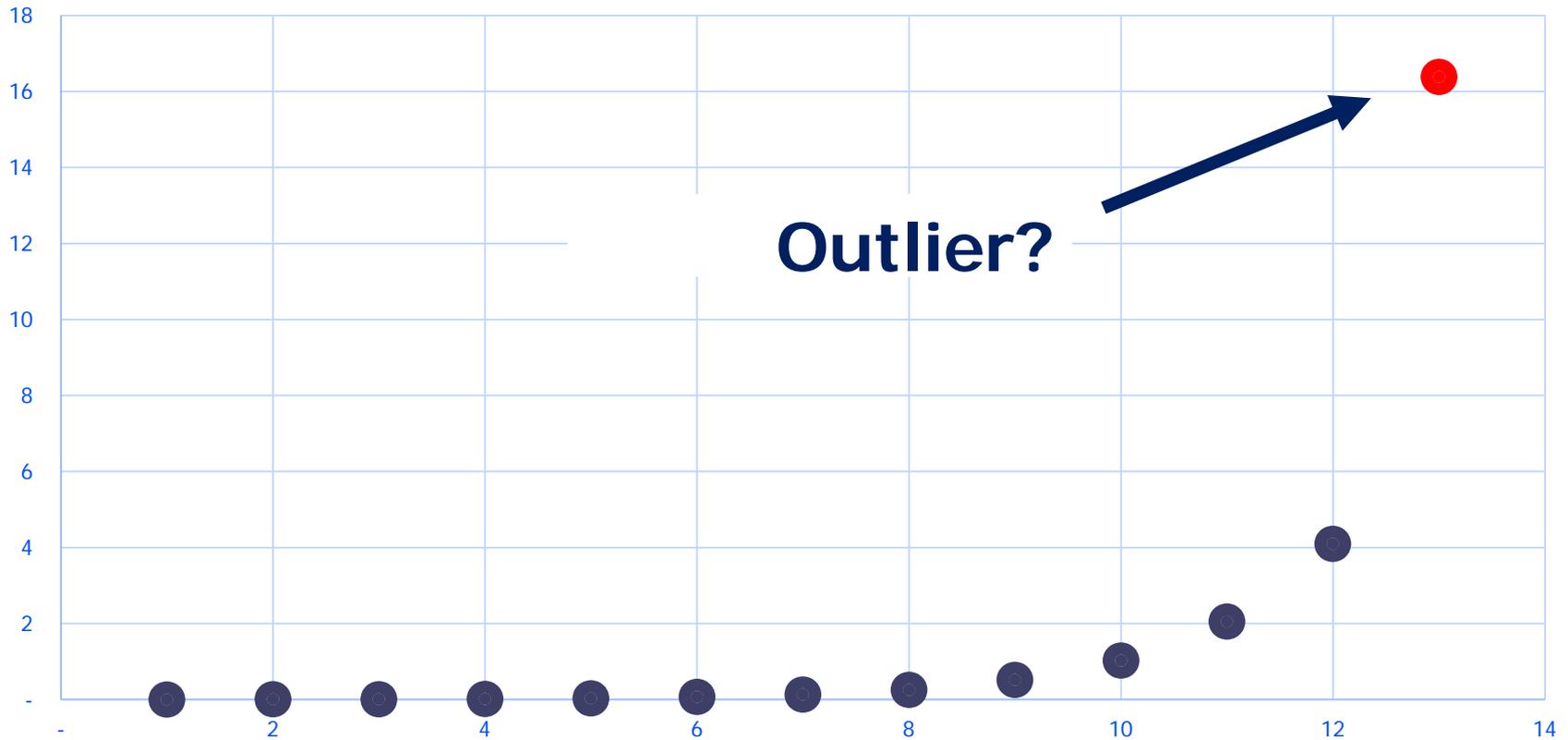
How to Determine Critical Values?

■ Percentiles:

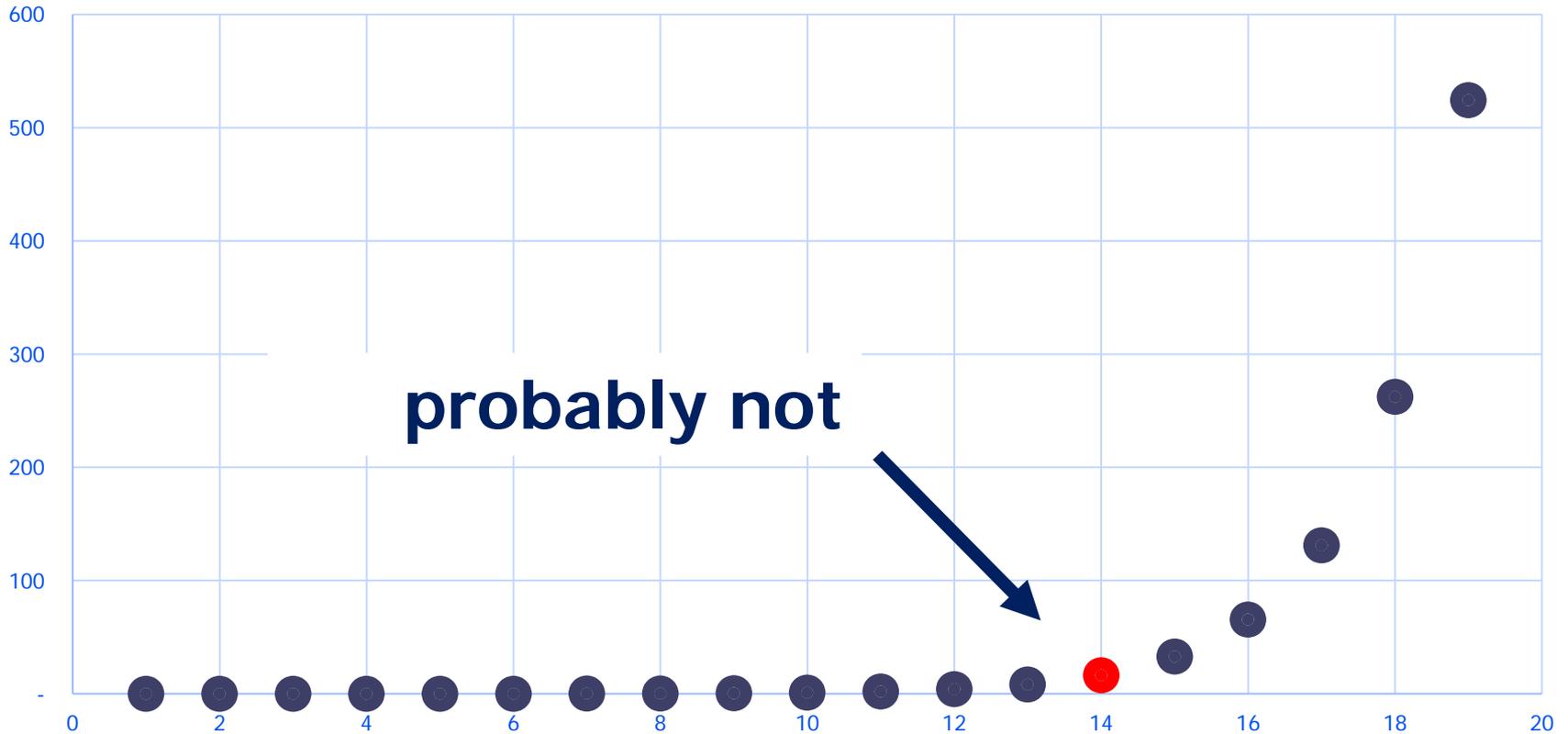
- ▶ Expenditures: 99.5 %
- ▶ Income: 97 %

■ Outside sources: If the sample differs from population

Distribution in Sample



Distribution in Population



How to Protect Respondents' Confidentiality?

- **Top-coding:** Provide average of expenditures above a threshold
- **Re-coding:** Change metadata but provide numerical data
- **Suppression:** Delete numerical data or entire record

How do We Recode?

- Find values that meet criteria
- Determine method:
 - ▶ Generalize info
 - ▶ Change info
- Replace original metadata with recoded metadata

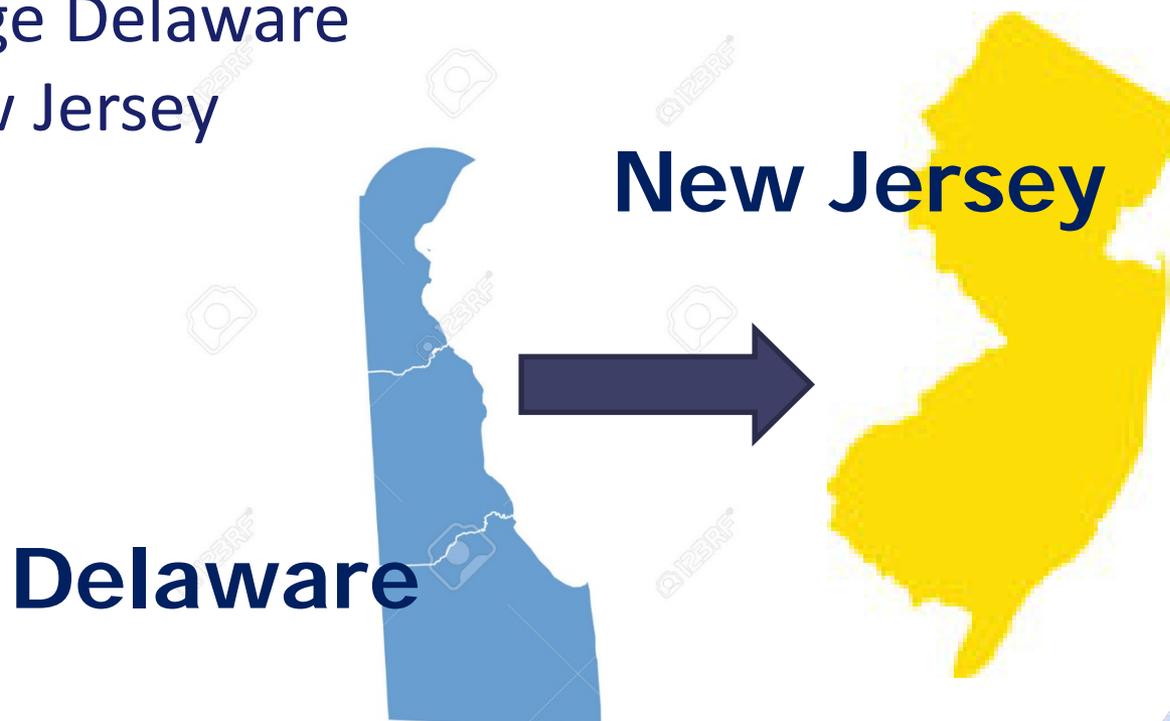
Re-coding: Generalize Information

- Broaden production year of cars
 - ▶ From Toyota Corolla 1999
 - ▶ To Toyota Corolla 1990s



Re-coding: Change information

- Change states to comparable states
 - ▶ Change Delaware to New Jersey



How to Protect Respondents' Confidentiality?

- **Top-coding:** Provide average of expenditures above a threshold
- **Re-coding:** Change metadata but provide numerical data
- **Suppression:** Delete numerical data or entire record

Suppression

Delete the reported data or
delete the entire record



How to Suppress?

- Erase numerical value
but maintain metadata
- Erase entire record
(Numerical value and metadata)

Suppression

- **Erase numerical data**
 - ▶ Blank values of normal but infrequent purchases
 - ▶ Example: Specialized mortgages

Suppression

- **Complete eradication**
 - ▶ Erase entire record
 - ▶ Example: Airplane purchase

Reverse Engineering

What's X?

$$5 = 3 + X$$

Reverse Engineering

Prevent the use of available information to deduce protected information

How to Prevent Reverse Engineering?

- Find protected values
- Protect them in all locations
- Protect related values

Reverse Engineering

- **Scenarios**
 - ▶ Within file
 - ▶ Across files

Reverse Engineering: Within File

■ Income = Wage + taxes

■ 1000 = 800 + 200

■ 1000 = 750 + 200

■ 950 = 750 + 200

■ Critical value: **700**

■ Topcode value: 750

Reverse Engineering: Within File

■ Income = Wage + taxes

■ 1000 = 800 + 200

■ 1000 = 750 + 200

■ 950 = 750 + 200

■ Critical value: 700

■ Topcode value: 750

Reverse Engineering: Within File

■ Income = Wage + taxes

■ 1000 = 800 + 200

■ **1000 = 750 + 200**

■ 950 = 750 + 200

■ Critical value: 700

■ Topcode value: 750

Reverse Engineering: Within File

■ Income = Wage + taxes

■ 1000 = 800 + 200

■ 1000 = 750 + 200

■ **950** = 750 + 200

■ Critical value: 700

■ Topcode value: 750

Reverse Engineering: Across Files

- **Income:** Topcoded income in FMLI
=> Topcode associated UCC in ITBI

- **Expenditure:** Topcoded expenditures in
EXPN/FMLI => topcode associated UCC in MTBI

How Do We Document?

■ Flag the values

- ▶ **T**: Topcoded value
- ▶ **D**: Valid value



Additional information

- [Protection of respondent confidentiality](#) provides more information on confidentiality
- See PUMD supplementary information page (http://www.bls.gov/cex/pumd_sup.htm)

Thank you!

Arcenis Rojas

Economist

Consumer Expenditure Survey

(202)-691-6884

Rojas.Arcenis@bls.gov