Overview of the Consumer Expenditure Surveys Program

Adam Safir, Division Chief

Consumer Expenditure Surveys Program

CE Microdata Users' Workshop

July 17, 2019

Topics

1. CE Program Overview
2. Annual Program Activities
3. CE Data Use

Bureau of Labor Statistics

Overview

- Principal fact-finding agency for the U.S. government in the field of labor economics
- Established in the Department of the Interior in 1884, to collect information about employment and labor; now part of Department of Labor

Core Values

- "Just the Facts"
- Gold standard data
- Transparency & confidentiality

CE Program Mission

- The mission of the CE is to collect, process, and disseminate information that presents a statistical picture of consumer spending for the Consumer Price Index, government agencies, and private data users.

- This mission encompasses analyzing CE data to produce socio-economic studies of consumer spending, and providing CE data users with assistance, education, and tools for working with the data.

CE Survey Overview

- The CE consists of estimates derived from two separate surveys
- The **Quarterly Interview Survey** is designed to collect data on large and recurring expenditures that consumers can be expected to recall for a period of 3 months or longer, such as rent and utilities (approximately 6,000 interviews/quarter)
- The **Diary Survey** is designed to collect data on small, frequently purchased items, including most food and clothing (approximately 3,000 diaries/quarter)
- Together, the data from the two surveys cover the complete range of consumers' expenditures

CE Survey Overview

- CE data are collected for BLS by the U.S. Census Bureau, from consumer units (CUs), in other words people living at one address who share living expenses (in most cases, CUs are the same as households)
- The sample frame is a national probability sample of households designed to be representative of the U.S. civilian noninstitutionalized population
- Respondents report dollar amounts for all non-investment purchases; business expenses and reimbursements are excluded

CE Program Staffing Structure

Division of Consumer Expenditure Surveys

- Branch of Information & Analysis
- Branch of Production & Control
- Branch of Research & Program Development

(incl. **Office of Survey Methods Research** consultants)

Division of Consumer Expenditure Information Systems

- Publication & Information Management Systems
- Expenditure & Income Processing Systems

Statistical Methods Division

- Consumer Expenditure Branch

CE Program Workflow

1. Sampling (SMD, P&C, Census)
2. Questionnaire Design and Revisions (P&C, Census)
3. Field Procedures and Training (P&C, Census)
4. Data Collection (P&C, Census)
5. Data Processing and Estimation (P&C, BIA, SMD, CEIS)
6. Data Dissemination, Publications, and Outreach (All)
7. Methods Research, Development, and Evaluation (BRPD, P&C, BIA, SMD)

CE Program Data Products

- CPI Cost Weights
- LABSTAT Database
- Publications
- **Public-Use Microdata**
- Restricted Microdata
- Tables

CE History – Highlights

- **1888:** First BLS expenditure survey
- **1972:** First use of weekly Diary & 3-month recall Interview
  Census Bureau begins survey collection for BLS
- **1979:** Continuous data collection starts
- **2003:** Computer Assisted Personal Interview (CAPI) begins
- **2004:** Imputation for missing income data implemented
- **2013:** Model-based estimation of income taxes introduced
- **2013:** 1st publication of midyear tables (July 2011 - June 2012)
- **2015:** Elimination of the first (bounding) interview in Interview
- **2018:** State-level weights released for 2016 and 2017 data

Topics

1. CE Program Overview
2. Annual Program Activities
3. CE Data Use

FY19 Program Activities

Data Collection

- Inputs for Interview and Diary Survey data collection protocol changes, training, and questionnaires

Data Processing

- Interview and Diary Survey processing (2018 & 2019)
- Interview and Diary Survey changes (2018 & 2019)

Data Dissemination

- Midyear Tables released in April 2019 (July 2017–June 2018 data)
- Annual Tables/Public-Use Microdata release in September 2019 (2018 data)

FY19 Program Activities

Publications

- Beyond the Numbers
- Data Comparisons (PCE, CPS, ACS)
- Monthly Labor Review
- Spotlight on Statistics

Outreach

- Conference Presentations
- Microdata Users' Workshop
- Survey Methods Symposium

FY19 Program Activities

Research

- Continue CE redesign activities, including development, and fielding, of the Large Scale Feasibility test of an online diary
- Investigate additional methods to evaluate and reduce measurement error in the CE Survey (e.g., a respondent worksheet)
- Work with Census to evaluate matched administrative data for potential production use in nonresponse adjustments

Topics

1. CE Program Overview
2. Annual Program Activities
3. CE Data Use

CE Customers

- Bureau of Economic Analysis
- Census Bureau
- Center for Medicare and Medicaid Services
- Consumer Price Index
- Depts of Agriculture; Defense; Health & Human Services
- Internal Revenue Service
- Media, Market Research, & Academic Research
- *YOU!*

CE Data in Recent External Publications

Generational Spending

- Why Do Millennials Hate Groceries?, The Atlantic

Regional Spending

- San Diego ranked No. 1 booziest city in America, FOX 5 San Diego

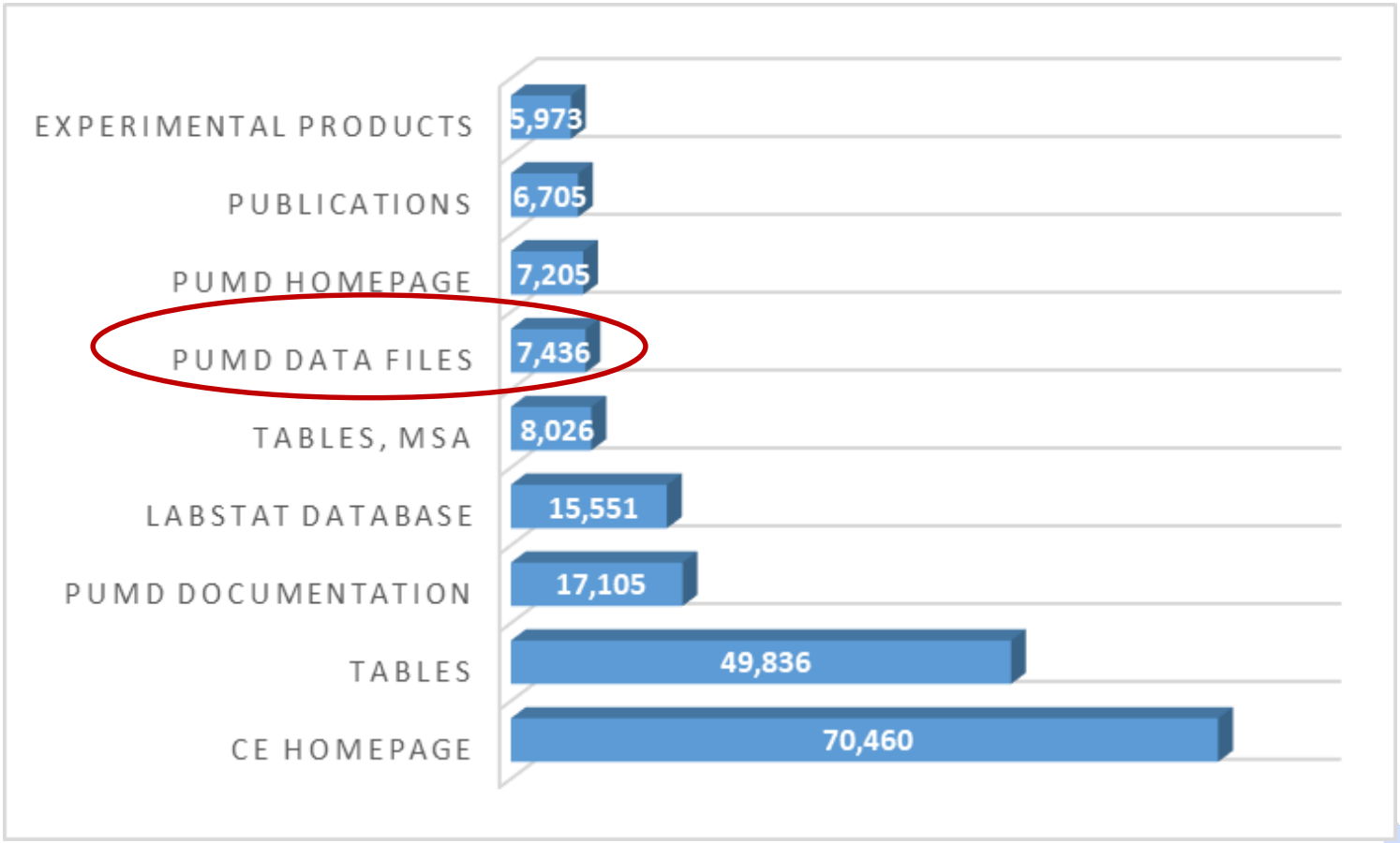
Health Care Spending

- The Financial Burden of Health Care Spending, Kaiser Family Foundation

Policy Analysis

- Tax Cuts And Jobs Act Begins To Have Positive Impact On Spending Intentions, Forbes
- Annual report on US consumption poverty: 2017 - AEI, AEI

Top 10 Webpages by Views, Jul-Dec 2018



Contact Information

Adam Safir
Consumer Expenditure Surveys Program
(202) 691-5175 | safir.adam@bls.gov

Consumer Expenditure Surveys Public-Use Microdata (PUMD) Overview

Jimmy Choi

Division of Consumer Expenditure Surveys
CE Microdata Users' Workshop
July 17-19, 2019

What Are Microdata?

- "In the study of survey and census data, microdata are information at the level of individual respondents." – Wikipedia
- Consumer Expenditure Surveys (CE) Public-Use Microdata (PUMD) consists of Consumer Unit (CU) level characteristics, income, taxes, assets and liabilities, and expenditure data, as well as member level characteristics and income data.
- Data are provided through various file types.


UNITED STATES DEPARTMENT OF LABOR
A to Z Index | FAQs | About BLS | Contact Us
Subscribe to E-mail Updates
GO
BUREAU OF LABOR STATISTICS
Follow Us | What's New | Release Calendar | Blog
Search BLS.gov
Home
Subjects
Data Tools
Publications
Economic Releases
Students
Beta
Consumer Expenditure Surveys
CE
SHARE ON:
PRINT:
BROWSE CE
CE HOME
CE OVERVIEW
CE NEWS RELEASES
CE PUBLICATIONS
CE TABLES
CE LABSTAT DATABASE
CE PUBLIC USE MICRODATA
CE WORKSHOP AND SYMPOSIUM
CE GEOGRAPHIC DATA
CE EXPERIMENTAL RESEARCH PRODUCTS
CONTACT CE
SEARCH CE
Go
CE TOPICS
INFORMATION FOR CE RESPONDENTS
CE LIBRARY
CE METHODOLOGY
CE SURVEY MATERIALS
CE FAQ
CE GLOSSARY
CE REDESIGN
CE DATA QUALITY AND
Public-Use Microdata (PUMD)
The Consumer Expenditure Survey (CE) collects data on expenditures, income, and demographics in the United States. The public-use microdata (PUMD) files provide this information for individual respondents, without any information that could identify respondents. PUMD files include adjustments for information that is missing because respondents were unwilling or unable to provide it. The files also have been adjusted to reduce the likelihood of identifying respondents, either directly or through inference.
This format allows researchers to analyze expenditures, income, and demographic trends beyond what the published tables show. However, users of the PUMD files should have strong skills with statistical software. To learn more, you may want to explore the PUMD methodology with the CE PUMD Getting Started Guide and check the availability of data with the Dictionary for Interview and Diary Surveys (XLSX).
The 2017 PUMD were released on September 11, 2018.
Note: To users of the CE 2017 Public-use Microdata: some corrected files have been reissued on June 27, 2019. See Errata notice for details.
Experimental State Weights for 2017, supporting state level analysis for selected states, are also available with their documentation.
To be notified when new datasets become available, please sign up for the CEX e-mail updates. If you have comments, suggestions, and questions about this page and its contents, please contact us.
https://www.bls.gov/cex/pumd.htm
Quick Links
» Introduction to PUMD
» Data files
» Documentation
» Supplemental resources
» Public-Use Microdata Annual Users' Workshop provides users the opportunity to learn how to use the PUMD and to present how they used PUMD in their research.

PUMD Data Package

- Data files:
  - File types available include:

  SAS (*.sas7bdat), STATA (*.dta), SPSS (*.sav), comma-delimited (*.csv)

- Documentation:
  - Documents covering specific years (changes with each new PUMD release)
  - Documents covering all years (Static or evolving documentation)

- Supplementary resources:
  - Provide information regarding PUMD disclosure methods and requirements

Consumer Expenditure Surveys

BROWSE CE
CE HOME
CE OVERVIEW
CE NEWS RELEASES
CE PUBLICATIONS
CE TABLES
CE LABSTAT DATABASE
CE PUBLIC USE MICRODATA
CE WORKSHOP AND SYMPOSIUM
CE GEOGRAPHIC DATA
CE EXPERIMENTAL RESEARCH PRODUCTS
CONTACT CE

SEARCH CE
Go

CE TOPICS
INFORMATION FOR CE RESPONDENTS
CE LIBRARY
CE METHODOLOGY
CE SURVEY MATERIALS
CE FAQ

PUMD Data Files

From 1996 forward, public-use microdata (PUMD) are available in four formats: SAS, SPSS, STATA and Comma Delimited (ASCII). Each format presents the data for each year in two zipped files. Data collected with the Interview Survey are in the "Interview" file and those with the Diary Survey are in the "Diary" file.

Pre-1996 PUMD are available as Comma Delimited files. The Inter-university Consortium for Political and Social Research (ICPSR) provides the data for free to its members. Here is their member list. BLS also sells the data on USB flash drives (PDF).

If you are new to the CE PUMD, you may want to explore the PUMD methodology with the CE PUMD Getting Started Guide and the availability of data with the Dictionary for Interview and Diary Surveys (XLSX). For those using income data in their analyses, we also recommend reading the User's Guide to Income Imputation in the CE. Additionally, we supply training programs and sample data (ZIP), and host a free microdata data users' workshop.

Experimental State Weights, supporting state-level analysis for selected states, are also available with their documentation.

The 2017 PUMD was released on September 11, 2018. To be notified when new datasets become available, please sign up for the CEX e-mail updates. If you have comments, suggestions, and questions about this page and its contents, please contact us.

If you are having difficulty downloading the data, please see the instructions on how to download zip files.

SAS | SPSS | STATA | CSV

https://www.bls.gov/cex/pumd_data.htm

SAS

2017 | Interview (zip) (89.1 MB) | Diary (zip) (10.6 MB) | Documentation
2016 | Interview (zip) (91.6 MB) | Diary (zip) (10.5 MB) | Documentation
2015 | Interview (zip) (87.7 MB) | Diary (zip) (10.5 MB) | Documentation

Data Files

- PUMD files contain respondent data collected in either the *Interview* or *Diary* survey.

- These data can be used to create and correlate information in a variety of ways:
  - Descriptive Statistics
    - How many CU's in the sample purchased a new car in 2014?
  - Inferential Statistics
    - What is the average annual expenditure on eggs by all CU's in the northeast region?


UNITED STATES DEPARTMENT OF LABOR
A to Z Index | FAQs | About BLS | Contact Us
Subscribe to E-mail Updates
GO
BUREAU OF LABOR STATISTICS
Follow Us | What's New | Release Calendar | Blog
Search BLS.gov
Home
Subjects
Data Tools
Publications
Economic Releases
Students
Beta
Consumer Expenditure Surveys
CE
SHARE ON:
PRINT:
BROWSE CE
CE HOME
CE OVERVIEW
CE NEWS RELEASES
CE PUBLICATIONS
CE TABLES
CE LABSTAT DATABASE
CE PUBLIC USE MICRODATA
CE WORKSHOP AND SYMPOSIUM
CE GEOGRAPHIC DATA
CE EXPERIMENTAL RESEARCH PRODUCTS
CONTACT CE
SEARCH CE
Go
CE TOPICS
INFORMATION FOR CE RESPONDENTS
CE LIBRARY
CE Experimental Research Products
This page provides an introduction to special tabulation tables and other experimental products being developed by the Consumer Expenditure Survey Division. New research products will be added and updated periodically.
Information on this page includes the All Consumer Unit Prepublication Table (All CU Prepublication MVP), new cross-tabulated tables with higher incomes, the 2014 Higher Income Table which expanded the top income range to $200,000 and over, a Generational Table that sorts expenditures by generation/cohorts, information on how to get the most from the CE published and prepublication tables, and the experimental state weight files.
Note that the 2014 through mid-2016 Generational tables were research work, and had not been produced using BLS production methods and standards. The 2016 annual Generational table is now included in the standard tables.
On This Page
» How to get the most from CE Tables
» NEW Cross-Tabulated Tables
» NEW State Weights
» Generational Research Tables
» All CU Prepublication MVP Tables
» 2014 Higher Income Table
https://www.bls.gov/cex/csxresearchtables.htm

Consumer Expenditure Surveys

BROWSE CE
CE HOME
CE OVERVIEW
CE NEWS RELEASES
CE PUBLICATIONS
CE TABLES
CE LABSTAT DATABASE
CE PUBLIC USE MICRODATA
CE WORKSHOP AND SYMPOSIUM
CE GEOGRAPHIC DATA
CE EXPERIMENTAL RESEARCH PRODUCTS
CONTACT CE

SEARCH CE
Go

CE TOPICS
INFORMATION FOR CE RESPONDENTS

PUMD Documentation

This page contains documentation for the public-use microdata (PUMD) for years starting in 1996. Documentation for years prior to 1996 are available USB flash drive for purchase (PDF). The documentation falls into two major types: Documentation that covers all years and documents that cover one particular year. If you are new to CE PUMD data, you may want to explore the CE PUMD Getting Started Guide.

Documents covering all years

- **Consumer Expenditure Surveys Public-use Microdata Getting Started Guide** provides documentation for the CE PUMD, its conventions, files, sample code, and methodology.
- **Consumer Expenditure Surveys Program Considerations When Using the Public-use Microdata** discusses considerations when preparing and interpreting estimates with PUMD.
- **Dictionary for Interview and Diary Surveys (XLSX)** provides variables and codes from 1996 forward.
- **Source selection file (XLSX)** identifies which survey data variable comes from when combining the two CE surveys for 1996 forward.
- **Description of income imputation** provides information on the methods BLS uses to estimate income since 2004.

Documents covering specific years

https://www.bls.gov/cex/pumd_doc.htm

- **Hierarchical groupings** lists the relation between the summary variables and their contributing variables as they are used in the published tables. Integrated stub (IntStub) lists the variables as BLS integrates them from the Interview and Diary Surveys

Documentation Covering All Years

- CE PUMD Getting Started Guide
  - Provides documentation for the CE PUMD, its conventions, files, sample code, and methodology.
- CE Program Considerations When Using the Public-use Microdata
  - Discusses considerations when preparing and interpreting estimates with PUMD

Documentation Covering All Years

- Dictionary for Interview and Diary Survey (XLSX)
  - Provides variables and codes from 1996 forward.
- Source selection file (XLSX)
  - Identifies which survey data item (category) comes from when combining the two CE surveys for 1996 forward.
- Description of Income Imputation
  - Provides information on the methods BLS uses to estimate income since 2004.

Documentation Covering Specific Years

- Hierarchical groupings
  - Establishes a relation between the published tables' series and their contributing expenditure classification codes (UCC).
- Sample code
  - Replicates totals in the CE published tables.
- Errata
  - Provides information on errors and their correction to the files of a particular year.

Consumer Expenditure Surveys

CE SHARE ON: PRINT:

BROWSE CE

- CE HOME
- CE OVERVIEW
- CE NEWS RELEASES
- CE PUBLICATIONS
- CE TABLES
- CE LABSTAT DATABASE
- CE PUBLIC USE MICRODATA
- CE WORKSHOP AND SYMPOSIUM
- CE GEOGRAPHIC DATA
- CE EXPERIMENTAL RESEARCH PRODUCTS
- CONTACT CE

SEARCH CE

Go

CE TOPICS

PUMD Supplementary Information

CE provides the following supplementary resources about the PUMD.

- **CE Glossary**
- **CE FAQ**
- **User guide to income imputation (PDF)**
- **Protection of respondent confidentiality** provides information on the methods the CE survey uses to protect the confidentiality of its respondents.

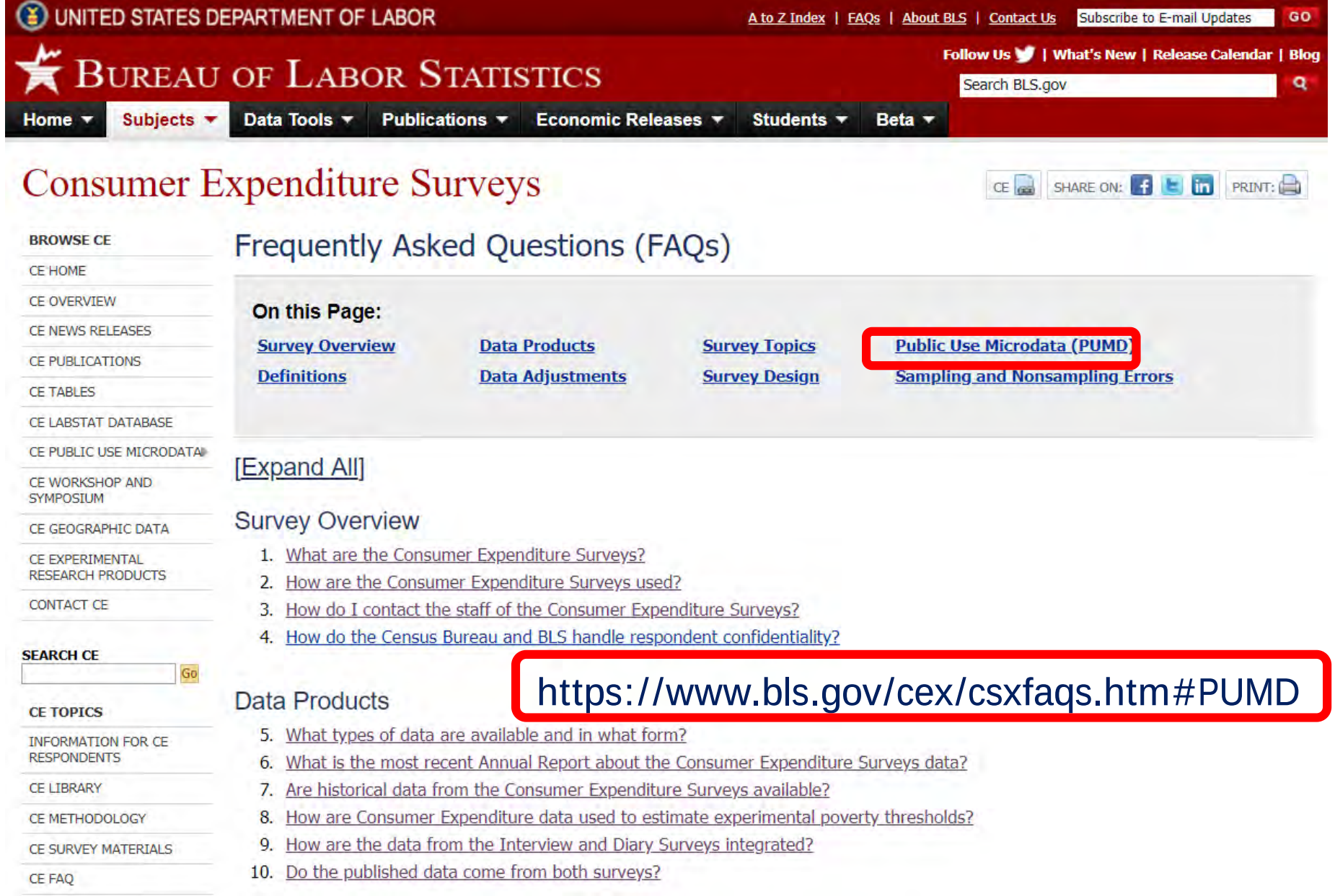
Last Modified Date: August 3, 2017

RECOMMEND THIS PAGE USING: Facebook Twitter LinkedIn

https://www.bls.gov/cex/pumd_sup.htm

Supplementary Resources

- CE Glossary
- CE FAQ
- User's guide to income imputation (PDF)
- Protection of respondent confidentiality


UNITED STATES DEPARTMENT OF LABOR
A to Z Index | FAQs | About BLS | Contact Us
Subscribe to E-mail Updates
GO
BUREAU OF LABOR STATISTICS
Follow Us | What's New | Release Calendar | Blog
Search BLS.gov
Home
Subjects
Data Tools
Publications
Economic Releases
Students
Beta
Consumer Expenditure Surveys
CE
SHARE ON:
PRINT:
BROWSE CE
CE HOME
CE OVERVIEW
CE NEWS RELEASES
CE PUBLICATIONS
CE TABLES
CE LABSTAT DATABASE
CE PUBLIC USE MICRODATA
CE WORKSHOP AND SYMPOSIUM
CE GEOGRAPHIC DATA
CE EXPERIMENTAL RESEARCH PRODUCTS
CONTACT CE
SEARCH CE
Go
CE TOPICS
INFORMATION FOR CE RESPONDENTS
CE LIBRARY
CE METHODOLOGY
CE SURVEY MATERIALS
CE FAQ
Frequently Asked Questions (FAQs)
On this Page:
Survey Overview
Data Products
Survey Topics
Public Use Microdata (PUMD)
Definitions
Data Adjustments
Survey Design
Sampling and Nonsampling Errors
[Expand All]
Survey Overview
1. What are the Consumer Expenditure Surveys?
2. How are the Consumer Expenditure Surveys used?
3. How do I contact the staff of the Consumer Expenditure Surveys?
4. How do the Census Bureau and BLS handle respondent confidentiality?
Data Products
5. What types of data are available and in what form?
6. What is the most recent Annual Report about the Consumer Expenditure Surveys data?
7. Are historical data from the Consumer Expenditure Surveys available?
8. How are Consumer Expenditure data used to estimate experimental poverty thresholds?
9. How are the data from the Interview and Diary Surveys integrated?
10. Do the published data come from both surveys?
https://www.bls.gov/cex/csxfaqs.htm#PUMD

2018 PUMD Release

- **September 10, 2019**

Jimmy Choi
Economist
Division of Consumer Expenditure Surveys
202-691-7081
Choi.Jimmy@bls.gov

Consumer Expenditure (CE) Public-Use Microdata (PUMD) File Structure

Taylor J. Wilson

Economist
Consumer Expenditure Surveys Division

CE Microdata User's Workshop
July 17 - 19, 2019

File Naming Conventions

- Naming conventions used for each PUMD release are common to both Interview and Diary data.
  - YY = Year
  - Q = Quarter
  - Q1 of 2017 → FMLI171 (2015 PUMD) or FMLI171x (2017 PUMD)
- The "X" signifies that the first quarter file of the current calendar year release is not identical to the fifth quarter file of the previous calendar year release.

The Interview Survey

Quarter 1 (fmli171x)

Oct 2016	Nov 2016	Dec 2016	Jan 2017	Feb 2017	March 2017
PQ	PQ	PQ	INT		
	PQ	PQ	CQ	INT	
		PQ	CQ	CQ	INT

Interview Quarterly Files

- FMLI - Characteristics, income, taxes, weights, and summary level expenditures for the Consumer Unit (CU) *as a whole*.
- MEMI - Characteristics and income for each *member in the CU*.
- MTBI - Detailed monthly expenditures.
- ITBI - Consumer Unit monthly income.
- ITII - Consumer Unit monthly imputed income.
- NTAXI - Federal and state tax information for each **tax unit** in the CU.

FMLI

Note: Not all the variables in the FMLI file are shown in the sample below

	NEWID	QINTRVMO	QINTRVYR	FAM_TYPE	REF_RACE	FINCBTXM	EDUCAPQ	EDUCACQ	FINLWT21	WTREP01
1	2649225	1	2014	1	1	105993.1	0	0	16314.976	NA
2	2659655	2	2014	8	1	32945.4	40	20	19811.63	38898.772
3	2707304	3	2014	7	6	78756	0	0	1996.748	NA
4	2710564	3	2014	1	1	383110	0	0	22809.279	43627.037
5	2758602	2	2014	7	5	42666	0	0	20405.795	NA
6	2699665	5	2014	1	1	79108.2	0	0	22256.228	42140.843
7	2729605	8	2014	1	1	13991	0	0	16790.071	NA
8	2816652	7	2014	3	4	225125.6	18000	0	20201.087	42959.282
9	2819892	7	2014	3	1	11173.8	120	0	14619.78	NA
10	2838742	8	2014	8	1	44600	0	0	18727.206	NA
11	2782324	10	2014	3	1	102000	0	0	15364.597	NA
12	2875873	12	2014	8	1	4400	0	0	15827.888	NA
13	2927262	11	2014	8	1	17123	0	0	23832.29	NA

MEMI

Note: Not all the variables in the MEMI file are shown in the sample below

	NEWID	MEMBNO	CU_CODE	AGE	EARNER	EDUCA	SEX	MEMBRACE	SALARYXM
1	2649225	1	1	65	2	4	2	1	NA
2	2649225	2	2	66	2	4	1	1	NA
3	2659655	1	1	46	1	7	2	1	32945.4
4	2816652	1	1	42	1	8	1	4	212937.2
5	2816652	2	2	39	2	8	2	1	NA
6	2816652	3	3	10	NA	NA	2	6	NA
7	2816652	4	3	8	NA	NA	2	6	NA
8	2816652	5	3	5	NA	NA	2	6	NA

MTBI

Note: Not all the variables in the MTBI file are shown in the sample below

	NEWID	EXPNAME	SEQNO	ALCNO	UCC	REF_MO	REF_YR	COST	COST_	PUBFLAG
1	2659655	JEDUCNET	3	0	670902	1	2014	20	D	2
2	2659655	JEDUCNET	3	0	670902	12	2013	20	D	2
3	2659655	JEDUCNET	3	0	670902	11	2013	20	D	2
4	2816652	JEDUCNET	27	0	670210	5	2014	18000	D	2
5	2816652	JEDUCNET	28	0	620310	4	2014	2800	D	2

ITBI

	NEWID	UCC	REFMO	REFYR	VALUE	VALUE_	PUBFLAG
1	2649225	900030	10	2013	1997.6	NA	2
2	2649225	900030	11	2013	1997.6	NA	2
3	2649225	900030	12	2013	1997.6	NA	2
4	2659655	5100	11	2013	18.9167	NA	2
5	2659655	5100	12	2013	18.9167	NA	2
6	2659655	5100	1	2014	18.9167	NA	2
7	2659655	900000	11	2013	2745.45	NA	2
8	2659655	900000	12	2013	2745.45	NA	2
9	2659655	900000	1	2014	2745.45	NA	2
10	2816652	900000	4	2014	17744.7704	T	2
11	2816652	900000	5	2014	17744.7704	T	2
12	2816652	900000	6	2014	17744.7704	T	2

ITII

	NEWID	UCC	REFMO	REFYR	IMPNUM	VALUE	VALUE_	PUBFLAG
1	2649225	900030	11	2013	1	2284	NA	2
2	2649225	900030	11	2013	2	2369	NA	2
3	2649225	900030	11	2013	3	2034	NA	2
4	2649225	900030	11	2013	4	2164	NA	2
5	2649225	900030	11	2013	5	1137	NA	2
6	2659655	900000	11	2013	1	2983.9167	NA	2
7	2659655	900000	11	2013	2	2785.4167	NA	2
8	2659655	900000	11	2013	3	2624.4167	NA	2
9	2659655	900000	11	2013	4	2741.9167	NA	2
10	2659655	900000	11	2013	5	2591.5833	NA	2
11	2816652	900000	4	2014	1	17744.7704	T	2
12	2816652	900000	4	2014	2	17744.7704	T	2
13	2816652	900000	4	2014	3	17744.7704	T	2
14	2816652	900000	4	2014	4	17744.7704	T	2
15	2816652	900000	4	2014	5	17744.7704	T	2

NTAXI

Note: Not all the variables in the MTBI file are shown in the sample below

	NEWID	TAX_UNIT	DEPCNT	FILESTAT	TAXYR_CY	TAXYR_PY	WAGE_HD	WAGE_SP	OTHTXINC	SOSSECB
1	2649225	1	0	2	2013	2012	0	0	0	23971
2	2659655	1	0	1	2013	2012	32945	0	0	0
3	2816652	1	3	2	2013	2012	212937	0	12188	0

Interview Annual Files

- EXPN – Expenditures by type (about 50 separate files)
  - Each file covers a class of expenditures such as vehicles and is derived from its corresponding section of the questionnaire.
- FPAR - Data about the survey.
- MCHI - Data about the contact history between the field representative and the respondent.

EDA (EXPN)

	NEWID	QYEAR	SEQNO	ALCNO	JEDUCNET	EDUC_AY	EDSCHL_A	EDMONTHA
1	2659655	20141	3	0	20	360	NA	13
2	2816652	20143	27	0	18000	300	2	5
3	2816652	20143	28	0	2800	100	NA	4

FPAR

Note: Not all the variables in the FPAR file are shown in the sample below

NEWID	QYEAR	OUTCOME	HOW_INTV	HSG_UNIT	INFOBOOK	LANGUAGE	RECORDS
2816652	20143	201	6	1	5	1	1
2649225	20141	201	6	1	5	1	4
2659655	20141	201	1	6	4	1	4

MCHI

Note: Not all the variables in the MCHI file are shown in the sample below

	NEWID	QYEAR	CNTCKEY	VISIT_YR	VISIT_MO	VISTWKDY	CTSTATUS	CTTYPE	INTERI
1	2649225	20141	506	2014	1	1	1	2	5
2	2649225	20141	507	2014	1	2	1	1	5
3	2659655	20141	501	2014	2	1	1	3	5
4	2659655	20141	506	2014	2	7	1	3	5
5	2659655	20141	507	2014	2	7	1	1	5
6	2816652	20143	205	2014	7	6	1	3	2
7	2816652	20143	208	2014	7	1	1	1	2

The Diary

Diary Files

- A Diary “quarter” refers to the calendar quarter in which the Diary booklet was placed in the home of the CU by the Census Field Representative.
- All Diary files are organized as quarterly files.
- A CU reports expenditures for two separate weeks.
- Each week’s diary is a separate record for each CU.
- Most Diary files are analogous to Interview files.

Diary Quarterly Files

- FMLD - a file with characteristics, income, weights, and summary level expenditures for the CU.
- MEMD - a file with characteristics and income for each member in the CU.
- EXPD - Detailed weekly expenditures.
- DTBD - Detailed annual income.
- DTID - Consumer Unit imputed income.

Two Ways to Organize the Data

FMLI

	NEWID	QINTRVMO	QINTRVYR	FAM_TYPE	REF_RACE	FINCBTXM	EDUCAPQ	EDUCACQ	FINLWT21	WTREP01
8	2816652	7	2014	3	4	225125.6	18000	0	20201.087	42959.282

Total Education Expenditures:
$18,000

FMLI

	NEWID	QINTRVMO	QINTRVYR	FAM_TYPE	REF_RACE	FINCBTXM	EDUCAPQ	EDUCACQ	FINLWT21	WTREP01
8	2816652	7	2014	3	4	225125.6	18000	0	20201.087	42959.282

EDA (EXPN)

	NEWID	QYEAR	SEQNO	ALCNO	JEDUCNET	EDUC_AY	EDSCHL_A	EDMONTHA
2	2816652	20143	27	0	18000	300	2	5
3	2816652	20143	28	0	2800	100	NA	4

Additional $2,800 ???

FMLI

	NEWID	QINTRVMO	QINTRVYR	FAM_TYPE	REF_RACE	FINCBTXM	EDUCAPQ	EDUCACQ	FINLWT21	WTREP01
8	2816652	7	2014	3	4	225125.6	18000	0	20201.087	42959.282

EDA (EXPN)

	NEWID	QYEAR	SEQNO	ALCNO	JEDUCNET	EDUC_AY	EDSCHL_A	EDMONTHA
2	2816652	20143	27	0	18000	300	2	5
3	2816652	20143	28	0	2800	100	NA	4

EDUC_AY:

Code 100 – Fees for recreational lessons

Code 300 – Tuition

From the Questionnaire

Section 16 - Educational Expenses

Section 16 collects educational expenses, including recreational lesson fees, tuition, room and board, purchases of school books and equipment, and other educational expenses. IMPORTANT: The Census Bureau does not release to the Bureau of Labor Statistics any confidential information such as names and addresses. This information is only used during the course of the interview.

Now I am going to ask about education expenses. Please include any direct payments made for any members of your household or for anyone outside your household and any payments you made online or had automatically deducted.

Do NOT include payments made on student loans.

1. Enter 1 to continue

For definitions Information Booklet »

Since the first of the reference month, have you or any members of your household paid for -

* Read each item on list
* Baby sitting and in home day care are collected in Section 19A.
* Do not include payments on student loans. They are collected in Section 22

1. Any recreational lessons or other instructions?
2. Preschool or child day care centers?
3. Tuition?
4. Housing while attending school?
5. Food or board while attending school?
6. Private school bus service?
7. Test preparation or tutoring services?
8. Purchase of any school books, supplies, or equipment which has not already been reported?
9. Other school related expenses not already reported?
99. None/No more entries

1. Recreational Lessons
3. Tuition

From the Dictionary

EDA – Detailed Expenditures Files (EXPN)

Educational Expenses

16 - Expenses paid by the CU

VARIABLE NAME	DESCRIPTION	FLAG	FORMAT	NOTE
EDUC_AY	Item code CODED 100 Recreational lessons or other instructions for members of this CU or other persons 200 Nursery school or child day care centers for members of this CU or other persons 300 Tuition 310 Housing while attending school 320 Food or board while attending school 330 Combined room and board (310, 320) 335 Combined tuition room and board (300, 310, 320) 340 Private school bus (previously captured in the XPB file under PRIVBUSX 345 Test Preparation or tutoring services 350 Purchase of any school books, supplies, or equipment which has not already been reported 360 Other school related expenses not already reported. (Now includes rental of any school books or equipment not already reported, previous code 340). 370 Combined expenses for books and tuition (300, 340- 350) 380 Other combined educational expenses (not previously reported) (100-320, 340-360) CAPI Section 16	EDUC_AY_	CHAR(3)	

From the Interview Stub File

Education Section

```
1  2    Education                                          EDUCATN     G  1  EXPEND
1  3      College tuition                                  670110      I  1  EXPEND
1  3      Elementary and high school tuition               670210      I  1  EXPEND
1  3      Vocational and technical school tuition          670410      I  1  EXPEND
1  3      Test preparation, tutoring services              670903      I  1  EXPEND
1  3      Other schools tuition                            670901      I  1  EXPEND
1  3      Other school expenses including rentals          670902      I  1  EXPEND
1  3      School books, supplies, equipment for college    660110      I  1  EXPEND
1  3      School books, supplies, equipment for elementary, high     660210      I  1  EXPEND
2           school
1  3      School books, supplies, equipment for vocational and       660410      I  1  EXPEND
2           technical schools
1  3      School books, supplies, equipment for day care, nursery    660901      I  1  EXPEND
1  3      School books, supplies, equipment for other schools        660902      I  1  EXPEND
```

Entertainment Section

```
1  2    Entertainment                                      ENTRTAIN    G  1  EXPEND
1  3      Fees and admissions                              FEESADM     G  1  EXPEND
1  4        Recreation expenses, out-of-town trips         610900      I  1  EXPEND
1  4        Social, recreation, health club membership     620111      I  1  EXPEND
1  4        Fees for participant sports                    620121      I  1  EXPEND
1  4        Participant sports, out-of-town trips          620122      I  1  EXPEND
1  4        Movie, theater, amusement parks, and other (thru           620211      I  1  EXPEND
2           Q20131)
1  4        Play, theater, opera, concert (new UCC Q20132)             620213      I  1  EXPEND
1  4        Movies, parks, museums (new UCC Q20132)        620214      I  1  EXPEND
1  4        Movie, other admissions, out-of-town trips     620212      I  1  EXPEND
1  4        Admission to sporting events                   620221      I  1  EXPEND
1  4        Admission to sports events, out-of-town trips  620222      I  1  EXPEND
1  4        Fees for recreational lessons                  620310      I  1  EXPEND
1  4        Other entertainment services, out-of-town trips 620903     I  1  EXPEND
```

FMLI

	NEWID	QINTRVMO	QINTRVYR	FAM_TYPE	REF_RACE	FINCBTXM	EDUCAPQ	EDUCACQ	FINLWT21	WTREP01
8	2816652	7	2014	3	4	225125.6	18000	0	20201.087	42959.282

EDA (EXPN)

	NEWID	QYEAR	SEQNO	ALCNO	JEDUCNET	EDUC_AY	EDSCHL_A	EDMONTHA
2	2816652	20143	27	0	18000	300	2	5
3	2816652	20143	28	0	2800	100	NA	4

MTBI

	NEWID	SEQNO	ALCNO	EXPNAME	COST	UCC	COST_	REF_MO	REF_YR	PUBFLAG
4	2816652	27	0	JEDUCNET	18000	670210	D	5	2014	2
5	2816652	28	0	JEDUCNET	2800	620310	D	4	2014	2

Organizing Expenditure Categories

- Interview EXPN files are organized according to the survey questionnaires.
  - Use the corresponding survey dictionary for all variables
- MTBI and Diary EXPD files are organized according to hierarchical groupings for publication
  - Use the corresponding stub file for UCC definitions
- FMLI and FMLD file summary variables aggregated using UCC groupings

Contact Information

Taylor J. Wilson
Economist
Division of Consumer Expenditure Surveys
202-691-6550
Wilson.Taylor@bls.gov

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

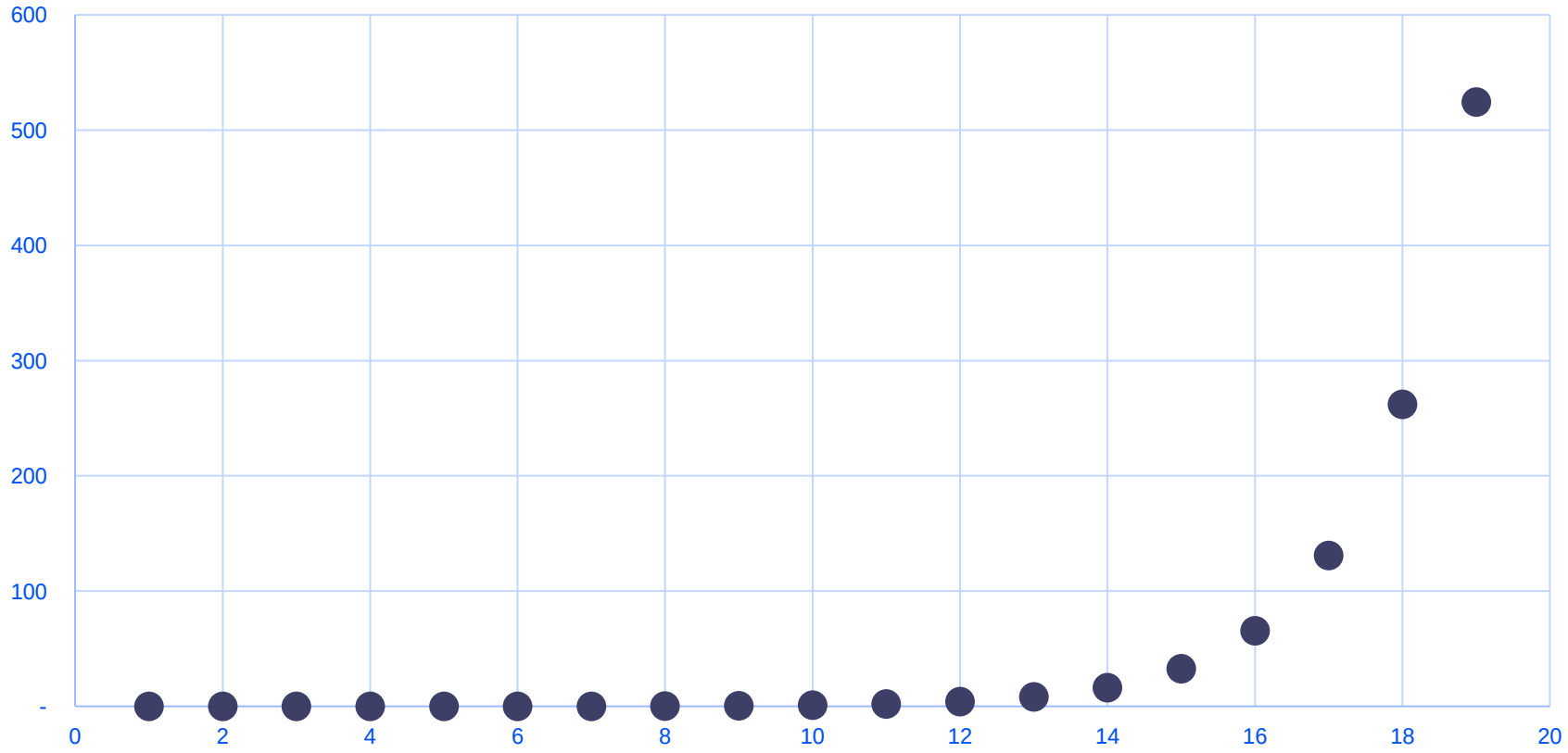
How to Conceal Indirect Identifiers?

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

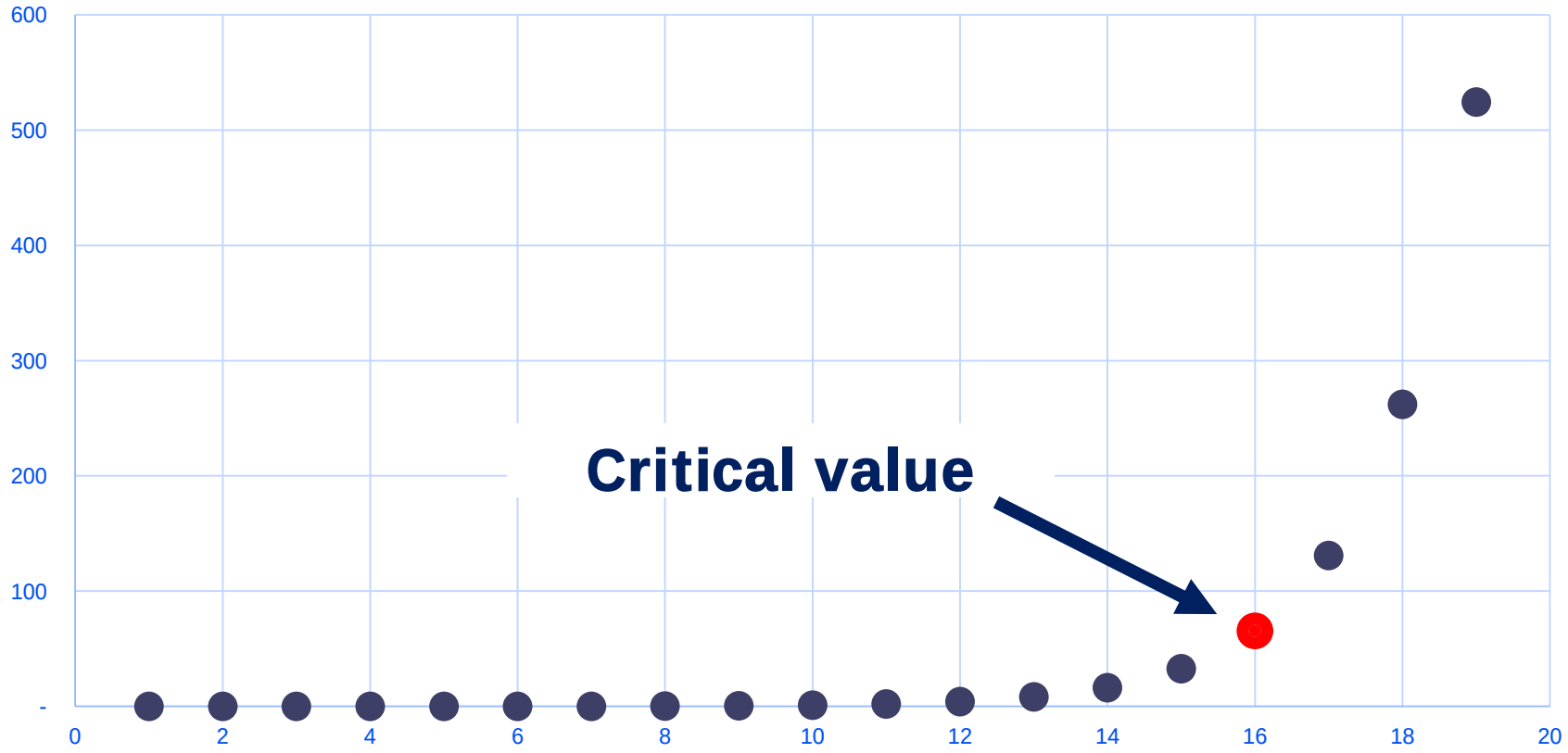
How do we Topcode?

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

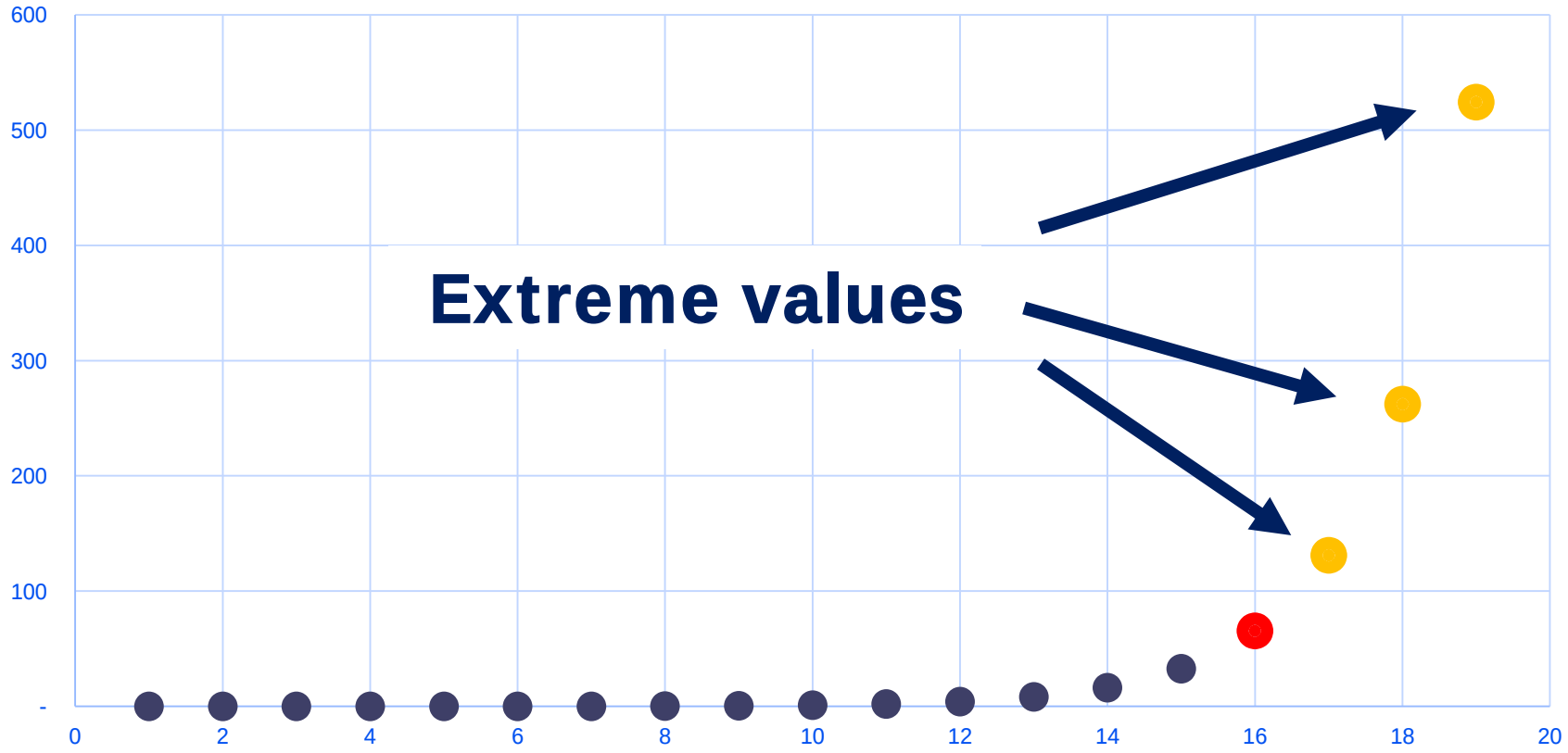
Topcoding Example



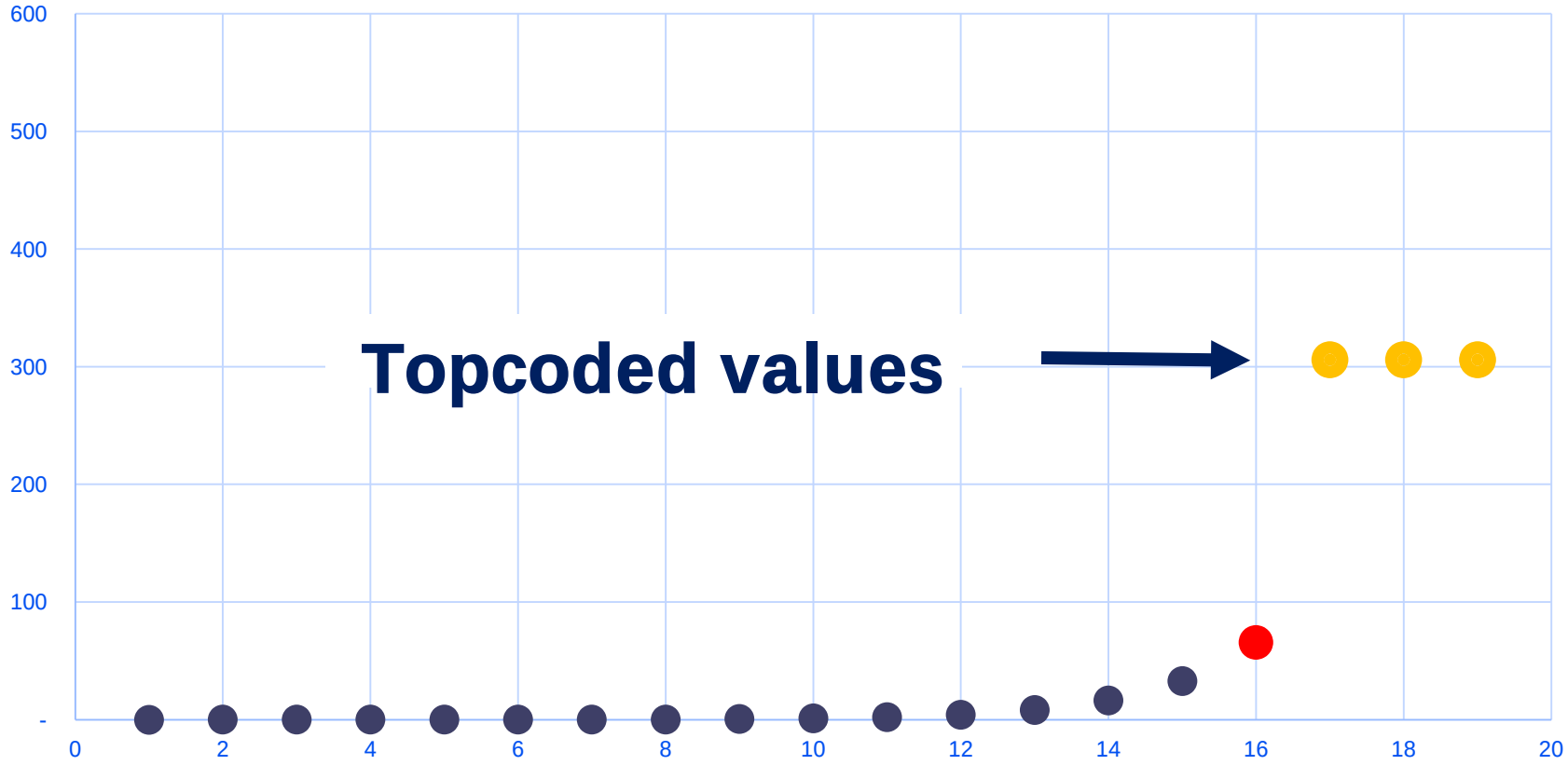
Topcoding Example



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?

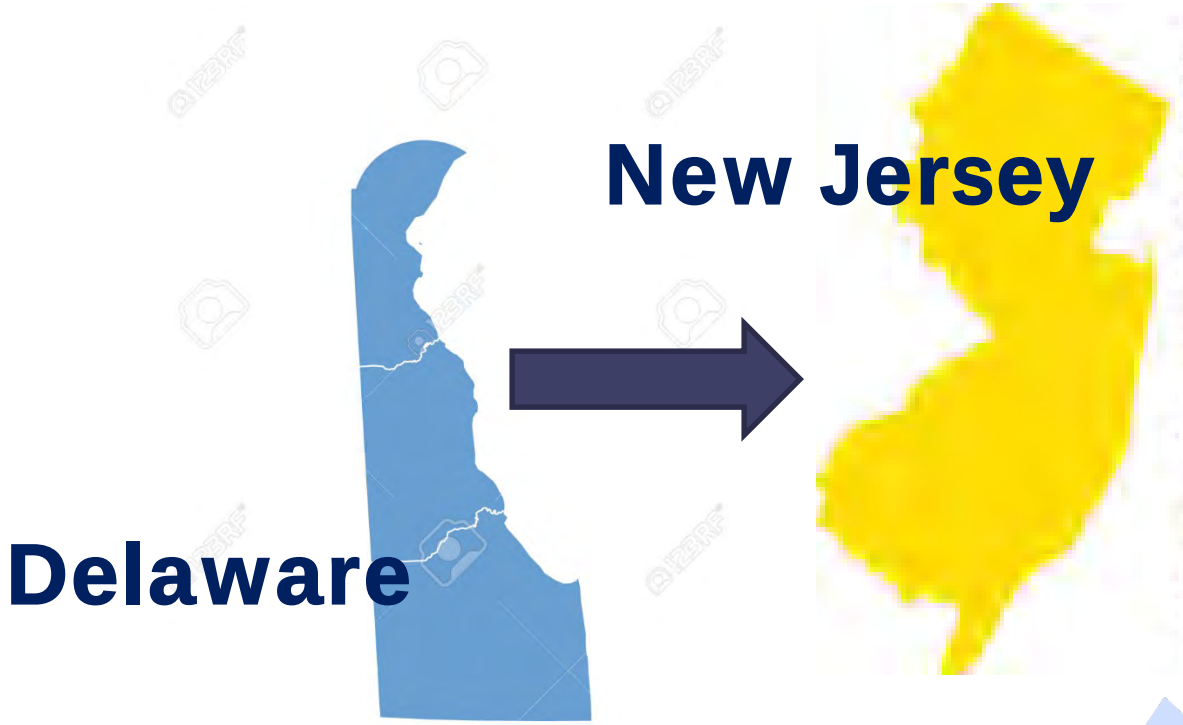
- Find revealing metadata
- Determine method:
  - Generalize information
  - Change information
- Replace original metadata with recoded metadata

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

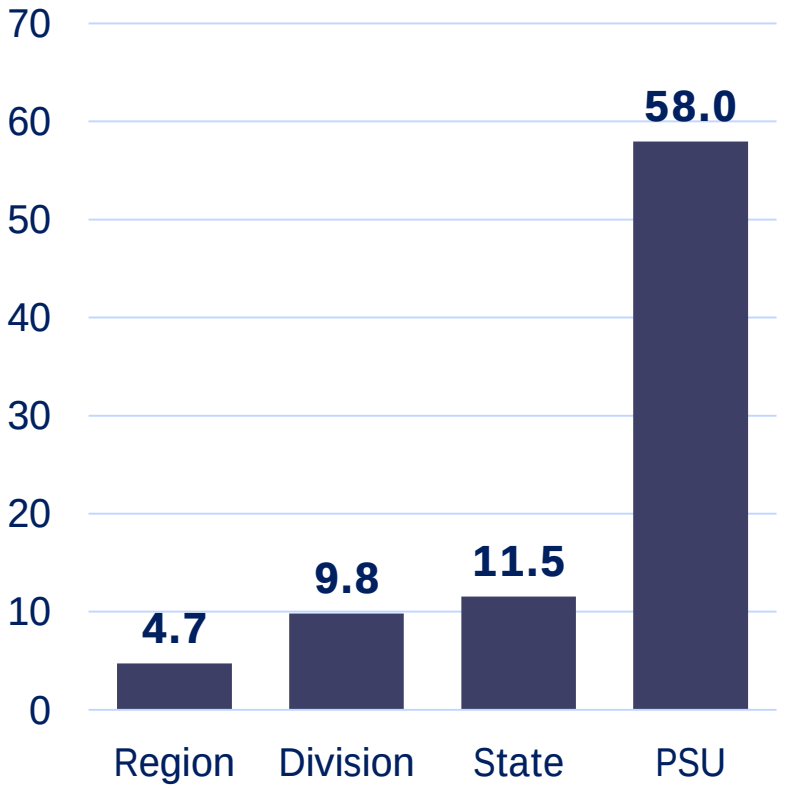
- **T**: Topcoded value
- **D**: Valid value, unadjusted

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

The CE Source Selection Process

Brett J. Creech
Branch of Information and Analysis
2019 Microdata Users Workshop
July 18, 2019

Overview

- Purpose
- Background
- Methodology
- Decision Criteria
- Next steps

What is Source Selection

- Methods used to select the appropriate survey for publication table estimates.
  - Interview
  - Diary
- For items that are unique to one survey or another, the choice is obvious.
- For items that overlap in coverage between the surveys, source selection methods are used to determine which source to select for publishing the Integrated data.

Purpose

- Primarily used for publication tables.
  - Identifies the more reliable source of survey data to use in estimation.

- PUMD data users
  - Provides a means for users to integrate survey estimates and closely replicate the publication tables.

Background

- The previous Source Selection method was developed in 1997 using data from 1993-1995. This method relied primarily on the Coefficient of Variance (CV) and in some cases the Mean Squared Error (MSE).
- In 2001, meetings were held with CE and CPI to look at differences in source selection using 1999 data. It was recommended that CPI adopt the CE source decision in all cases with greater than 50 reports of expenditures at the UCC level.

Background

- In 2006, when incorporating a few new UCCs with 2005 data, source selection was coordinated so that CE and CPI were in agreement on the newly introduced UCCs.

- In 2007, CE and CPI formed a team to evaluate and come up with a new methodology for Source Selection to be used for 2007 publication.

Methodology – Overall Goal

- Over 200 UCCs are processed using both the Diary and the Interview.
- A determination is made to which source is used for the integrated tables.
- The overall methodology selects the higher mean given two decision criteria with exceptions from the CPI.

Methodology

- Preliminary steps:
  - Calculating counts, sample means, and sample variances.
  - Data are top coded and bottom coded
    - This is done to minimize the impact of outliers

Methodology

- The counts (representing a reported expenditure for that UCC) and Z-scores are weighted for the three most recent collection years:
  - 1st collection year by 1/6 (For the 2017 data, use 2014)
  - 2nd collection year by 2/6 (For the 2017 data, use 2015)
  - 3rd collection year by 3/6 (For the 2017 data, use 2016)

Methodology

- If a new UCC was created in the past 2 years (for example, a new UCC created in 2015), then the following weights are used:
  - 1st collection year by 2/5 (For 2017, use 2015 data)
  - 2nd collection year by 3/5 (For 2017, use 2016 data)

Decision Criteria

- There are two criterions that are used in determining source selection decisions:
  - Criterion 1: Counts Sufficiency
  - Criterion 2: Statistical Significance

Criterion 1: Counts Sufficiency

- For each UCC and each survey (Interview or Diary), the number of consumer units with at least one expenditure is counted for each of the 3 most recent data years.
  - Yields 6 counts for each UCC
    - Three yearly counts for Interview
    - Three yearly counts for Diary

Criterion 1: Counts Sufficiency

- A sufficient amount of data exists when the count for each of the 3 years is greater than or equal to 60.
- If both surveys have sufficient data then proceed to the next Criterion.
- If both surveys lack sufficient data, then keep existing source.

Criterion 1: Counts Sufficiency

- If one survey has sufficient data, but the other has insufficient data, then a weighted average of the three yearly counts for the survey having an insufficient amount of data is computed: $n^* = (3/6)n_{t-1} + (2/6)n_{t-2} + (1/6)n_{t-3}$

Criterion 1: Counts Sufficiency

- If the weighted average n* from the insufficient survey is greater than or equal to 60, then proceed to the next Criterion.

- If the weighted average n* from the insufficient survey is still less than 60, then use the survey with sufficient data as the source.

Criterion 2: Statistical Significance (Z-scores)

- If the value of the weighted Z-Score, $z^* = (3/6)z_{t-1} + (2/6)z_{t-2} + (1/6)z_{t-3}$, is greater than or equal to 1.645, or less than or equal to -1.645 then select the source based on the following:
  - Greater than or equal to 1.645 – Interview Survey
  - Less than or equal to -1.645 – Diary Survey

Criterion 2: Statistical Significance (Z-scores)

- If the weighted Z-Score is between –1.000 and 1.000, then the current source will continue to be used.

Criterion 2: Statistical Significance (Z-scores)

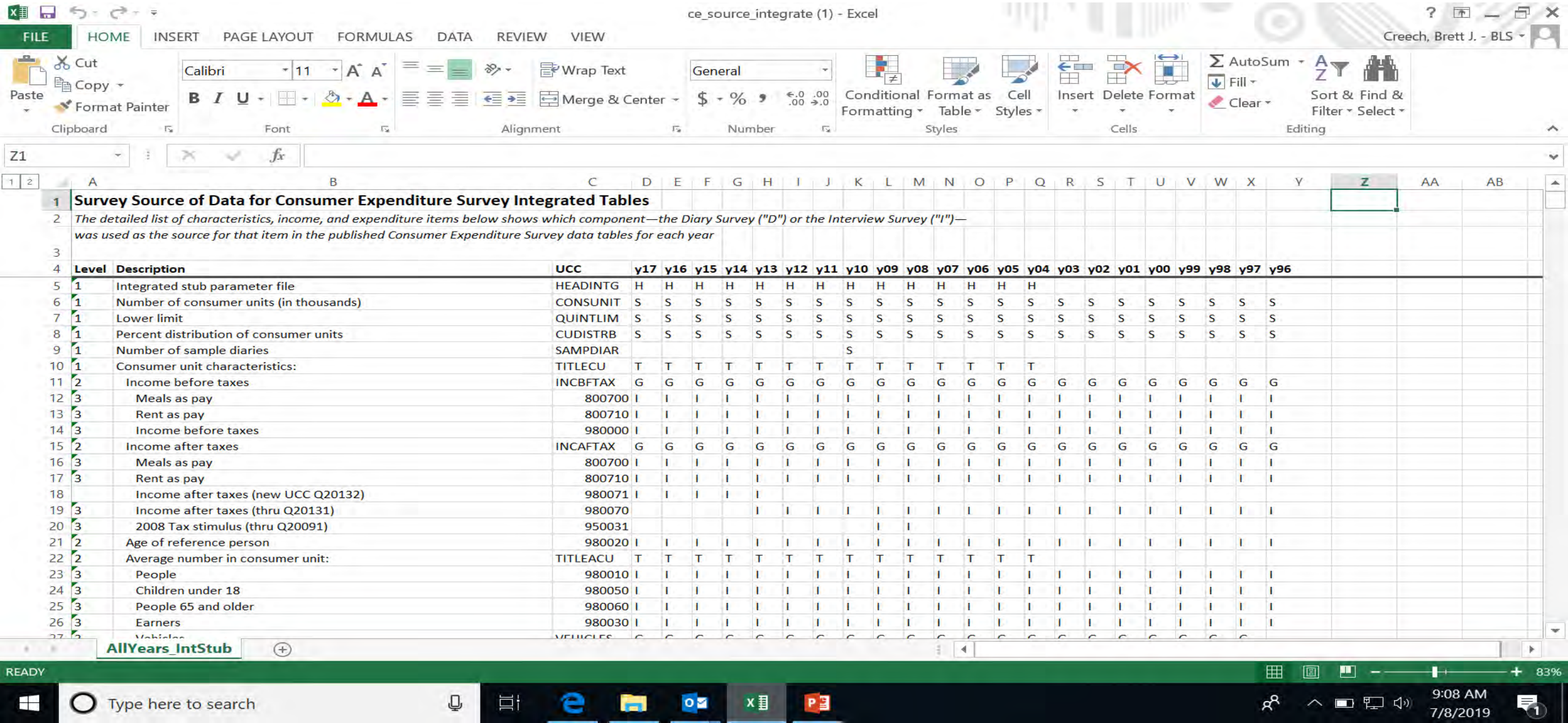
- If all three z-scores are 1.000 and above, then use the Interview Survey
- If all three z-scores are -1.000 and lower, then use the Diary Survey
- Any remaining scenarios, the source remains the same.

Exclusions – Items stay in the Interview Survey

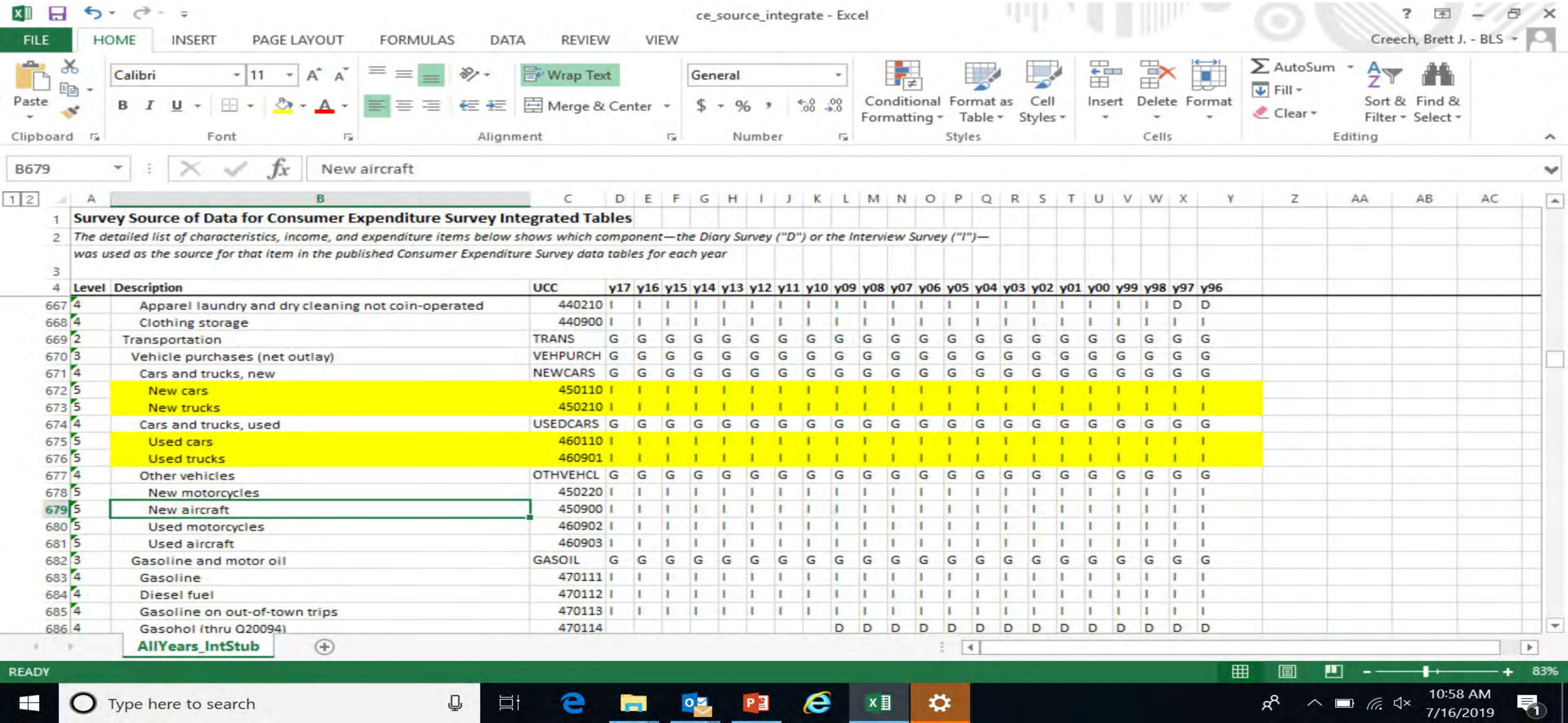
- Expenditures for items net of reimbursements
  - Medical Care
  - Auto Repairs
- Reimbursements are captured in the Interview survey
  - Not captured in the Diary survey
- Transportation UCCs
  - Trade-in vehicle values are deducted from purchases of new cars in out-of-pocket expense calculations

Where to find the Source Selection spreadsheet

Spreadsheet 1996-2017

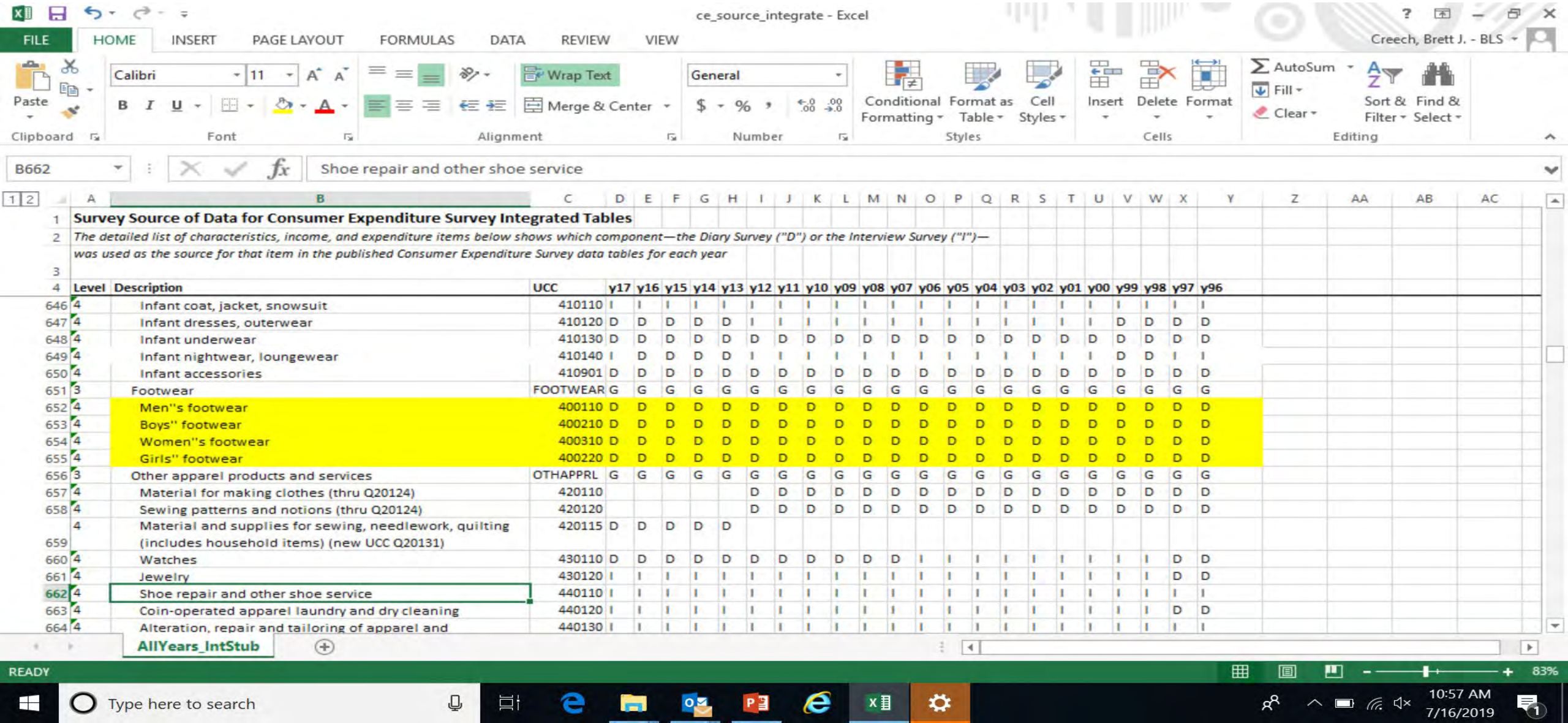


Interview Example

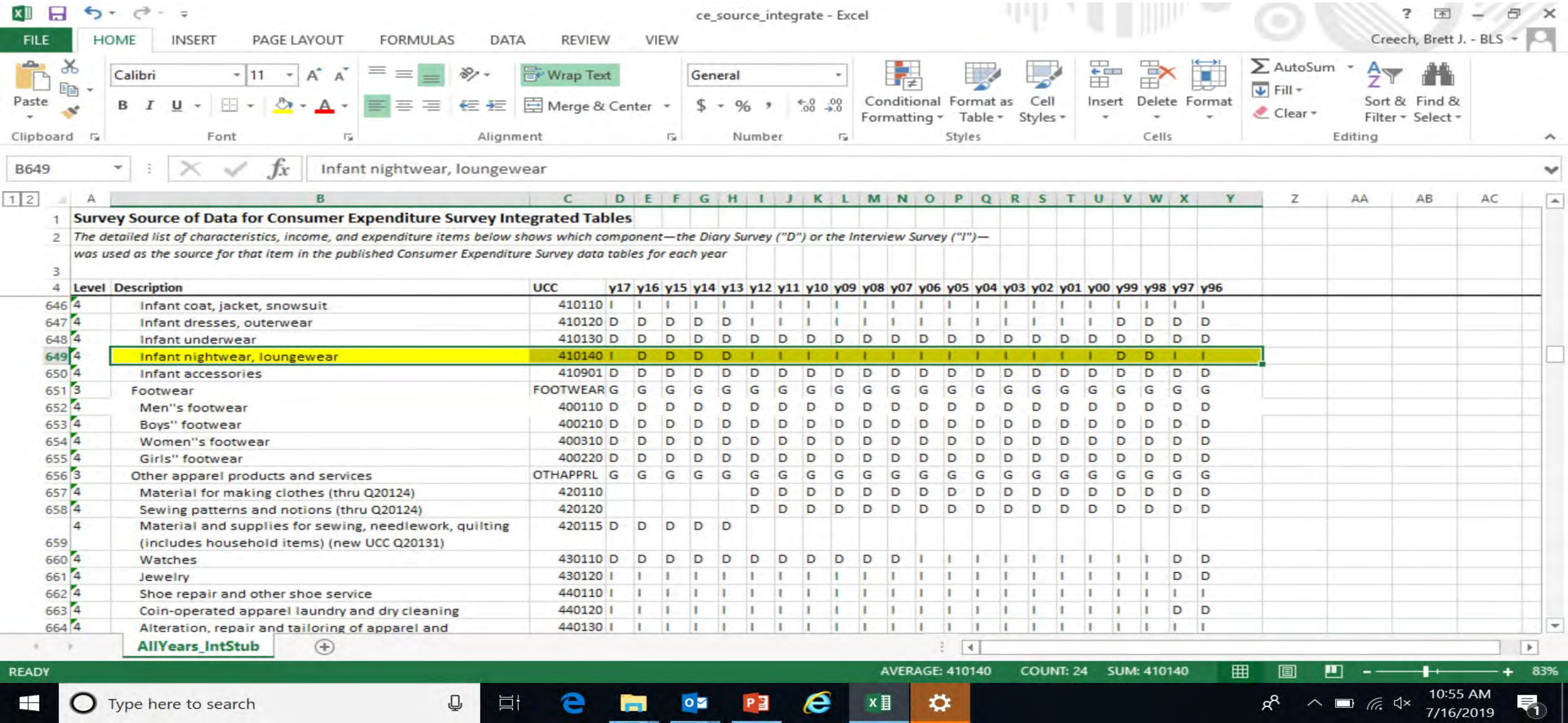


	Level	Description	UCC	y17	y16	y15	y14	y13	y12	y11	y10	y09	y08	y07	y06	y05	y04	y03	y02	y01	y00	y99	y98	y97	y96
667	4	Apparel laundry and dry cleaning not coin-operated	440210	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	D	D
668	4	Clothing storage	440900	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
669	2	Transportation	TRANS	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
670	3	Vehicle purchases (net outlay)	VEHPURCH	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
671	4	Cars and trucks, new	NEWCARS	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
672	5	New cars	450110	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
673	5	New trucks	450210	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
674	4	Cars and trucks, used	USEDCARS	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
675	5	Used cars	460110	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
676	5	Used trucks	460901	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
677	4	Other vehicles	OTHVEHCL	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
678	5	New motorcycles	450220	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
679	5	New aircraft	450900	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
680	5	Used motorcycles	460902	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
681	5	Used aircraft	460903	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
682	3	Gasoline and motor oil	GASOIL	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
683	4	Gasoline	470111	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
684	4	Diesel fuel	470112	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
685	4	Gasoline on out-of-town trips	470113	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
686	4	Gasohol (thru Q20094)	470114									D	D	D	D	D	D	D	D	D	D	D	D	D	D

Diary Example



Source Selection Change



Survey Source of Data for Consumer Expenditure Survey Integrated Tables

The detailed list of characteristics, income, and expenditure items below shows which component—the Diary Survey ("D") or the Interview Survey ("I")—was used as the source for that item in the published Consumer Expenditure Survey data tables for each year

	Level	Description	UCC	y17	y16	y15	y14	y13	y12	y11	y10	y09	y08	y07	y06	y05	y04	y03	y02	y01	y00	y99	y98	y97	y96
646	4	Infant coat, jacket, snowsuit	410110	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
647	4	Infant dresses, outerwear	410120	D	D	D	D	D	I	I	I	I	I	I	I	I	I	I	I	I	I	D	D	D	D
648	4	Infant underwear	410130	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
649	4	Infant nightwear, loungewear	410140	I	D	D	D	D	I	I	I	I	I	I	I	I	I	I	I	I	I	D	D	I	I
650	4	Infant accessories	410901	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
651	3	Footwear	FOOTWEAR	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
652	4	Men''s footwear	400110	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
653	4	Boys'' footwear	400210	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
654	4	Women''s footwear	400310	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
655	4	Girls'' footwear	400220	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
656	3	Other apparel products and services	OTHAPPRL	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
657	4	Material for making clothes (thru Q20124)	420110						D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
658	4	Sewing patterns and notions (thru Q20124)	420120						D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
659	4	Material and supplies for sewing, needlework, quilting (includes household items) (new UCC Q20131)	420115	D	D	D	D	D																	
660	4	Watches	430110	D	D	D	D	D	D	D	D	D	D	D	I	I	I	I	I	I	I	I	I	D	D
661	4	Jewelry	430120	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	D	D
662	4	Shoe repair and other shoe service	440110	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
663	4	Coin-operated apparel laundry and dry cleaning	440120	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	D	D
664	4	Alteration, repair and tailoring of apparel and	440130	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Reference

Brett Creech and Barry Steinberg: CE Source Selection for Publication Tables

https://www.bls.gov/cex/anthology11/csxanth3.pdf

What's next?

- Team is being formed to revisit the current Source Selection methodology
- 2019 data: Use current methodology while testing new approach
- 2021 data: Potential new methodology implemented

Contact Information

Brett J. Creech
(202) 691-5120
Creech.Brett@bls.gov

2019 CE Survey Microdata Users' Workshop

Sampling Methods and Derivation of Sampling Weights

Brian T. Nix

Consumer Expenditure Survey
Statistical Methods Division
Bureau of Labor Statistics

July 18, 2019

Overview

- History and Concepts
- Sample Selection
  - Define PSUs
  - Stratify and Select a Sample of PSUs
  - Stratify and Select a Sample of Households
- Weighting the Households

History of Sample Redesigns

- New sample of geographic areas selected every decade
  - **1980 Census-Based Sample Design (1986–1995)**
  - **1990 Census-Based Sample Design (1996–2004)**
  - **2000 Census-Based Sample Design (2005–2014)**
  - **2010 Census-Based Sample Design (2015–2024)**
  - **2020 Census-Based Sample Design (2025–2034???)**

Concepts

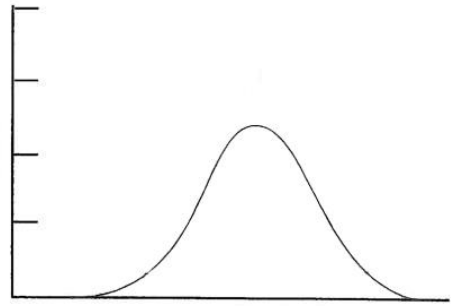
- Target Population
  - U.S. non-institutional civilian population
- PSU (Primary Sampling Unit)
  - Geographic area used for sampling
  - Cluster of contiguous counties
    - (between 2 and 5 counties on average)
- CU (Consumer Unit)

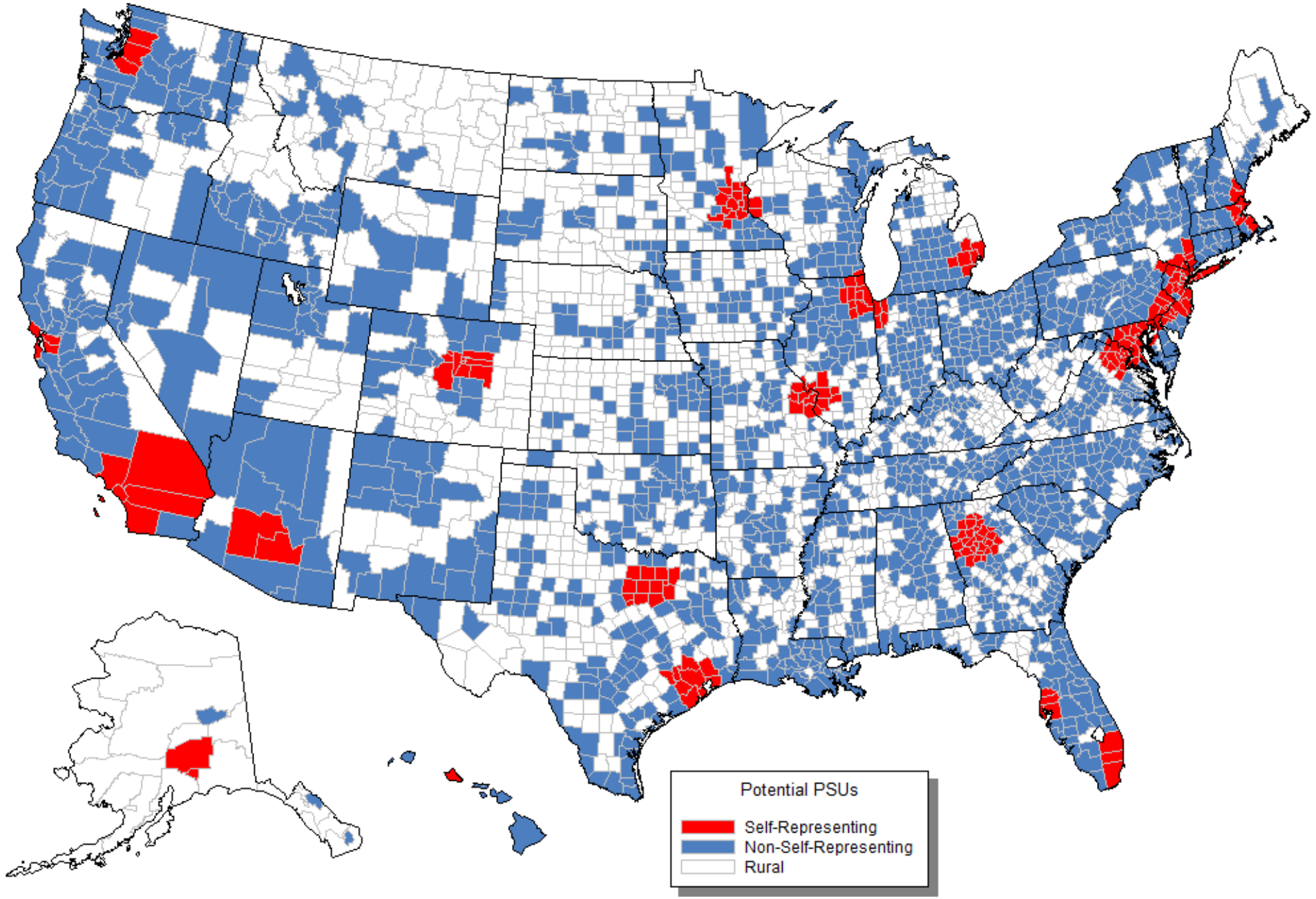
≈ Household

Sample Selection – Overview

- Geographic areas are randomly selected to represent the total U.S.
- Households are randomly selected to represent the geographic areas
- Guiding principle:

  "***Randomness ensures representativeness.***"

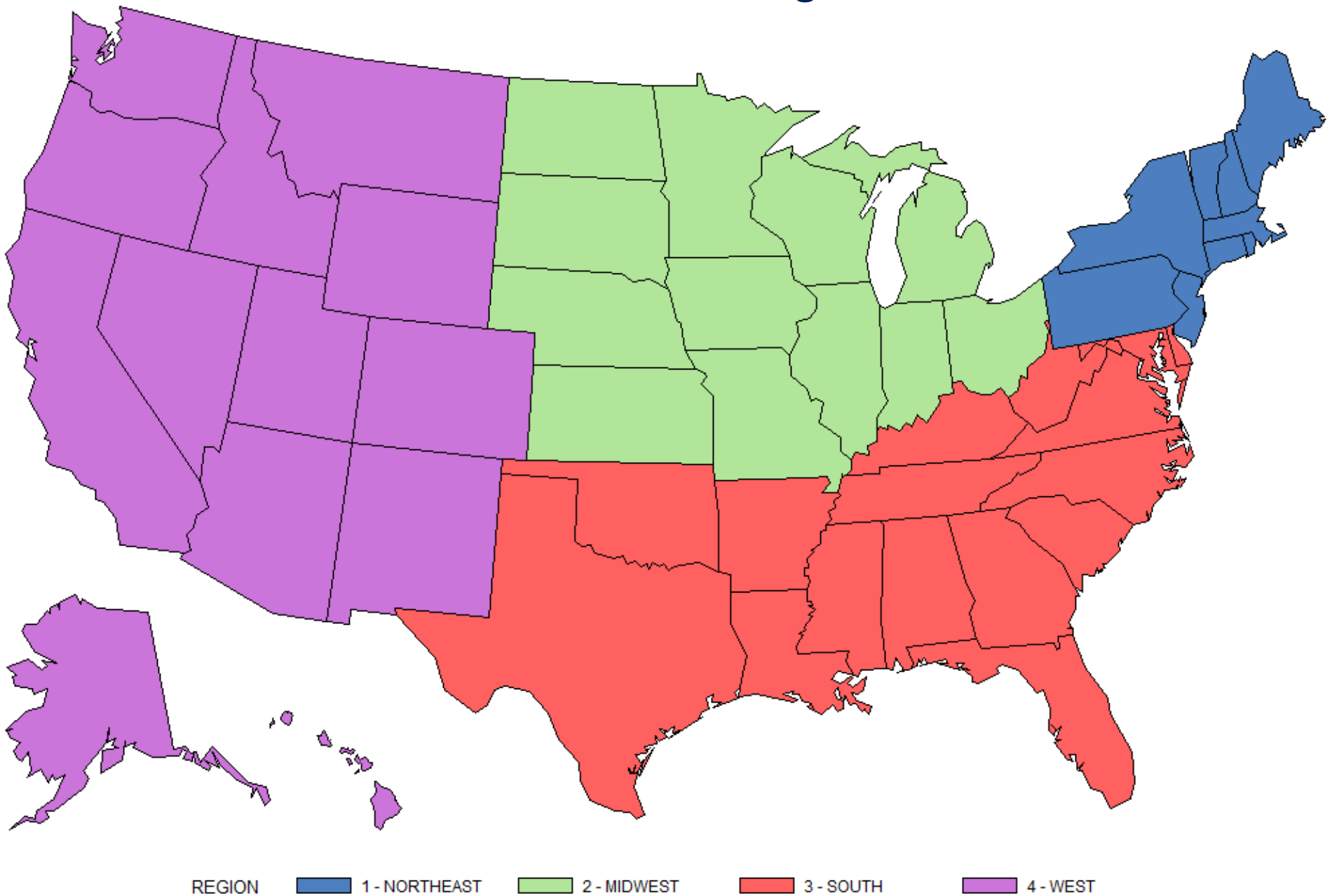



Potential PSUs
Self-Representing
Non-Self-Representing
Rural

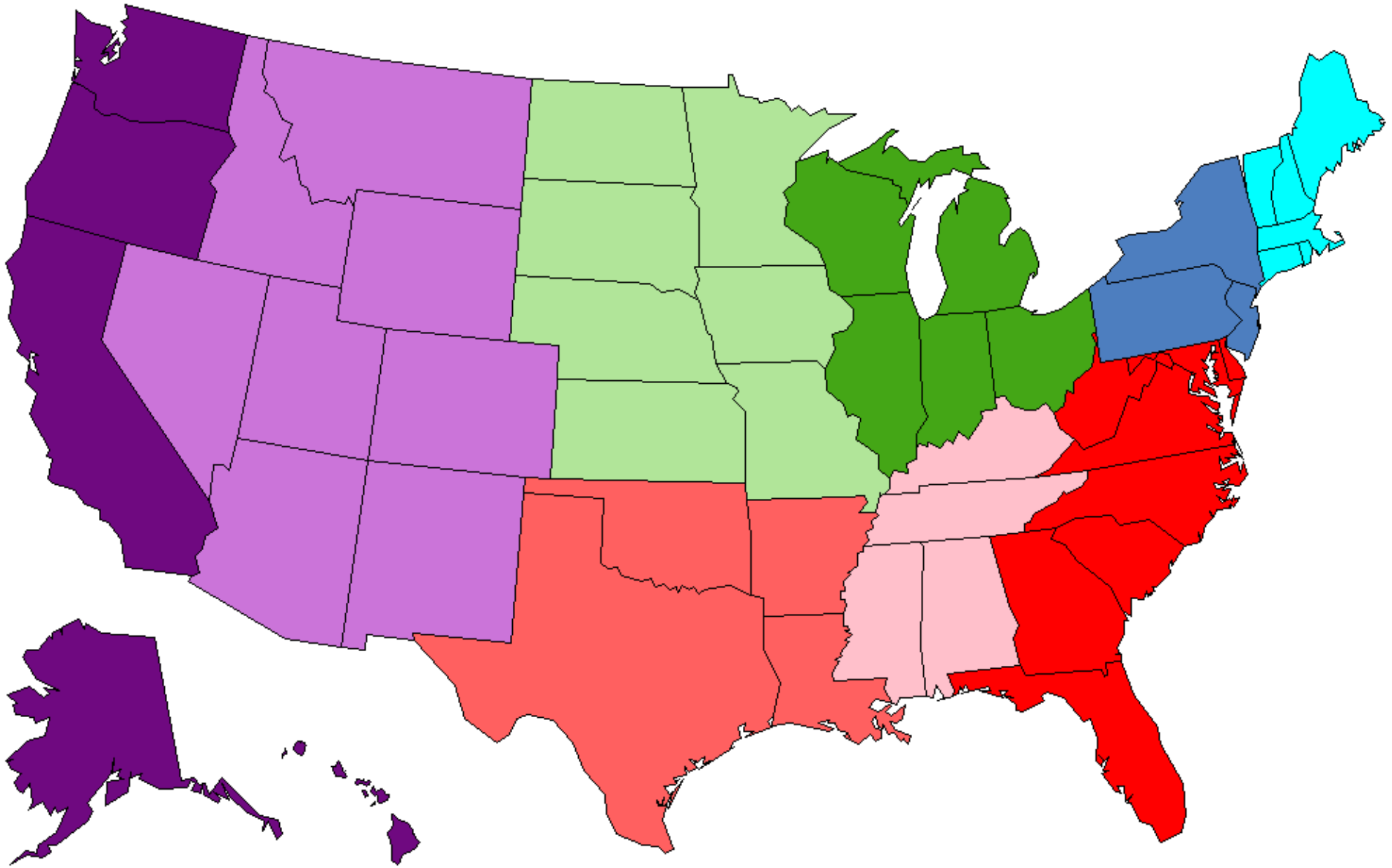
Selection of PSUs

PSU class	Description	CBSA/ Non-CBSA	Population Total	Examples
S	Self-Representing	Metropolitan (urban)	Greater Than or Equal to 2,500,000	S11A Boston MA S49D Seattle WA
N	Non-Self-Representing	Metro- or Micropolitan (urban)	Less Than 2,500,000	*Suppressed*
R	Rural *(also not Self-Representing)*	Non-CBSA (rural)		*Suppressed*

The Four Census Regions



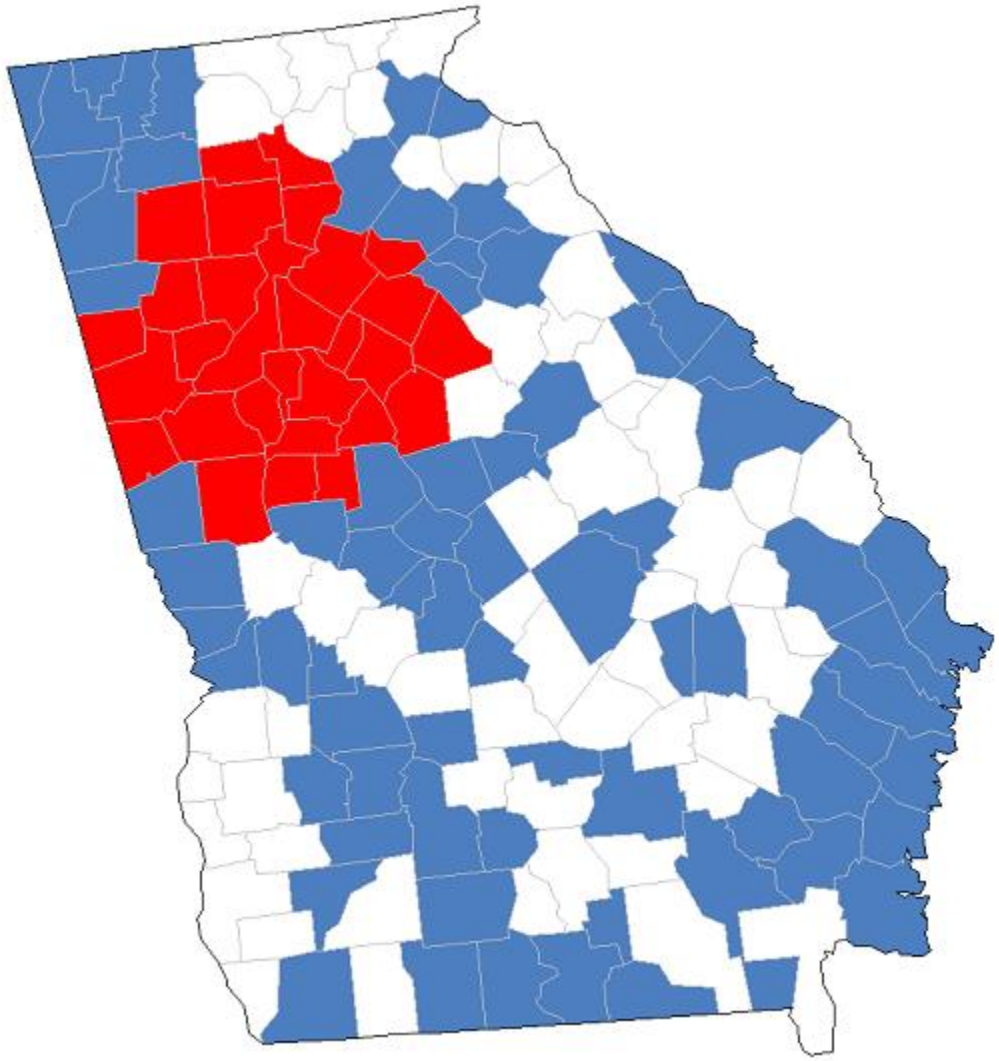
The Nine Census Divisions



Sample Selection:
CPI – 75 PSUs; CE – 91 PSUs

PSU Size	Region/Division									Total
	Northeast		Midwest		South			West		
	01	02	03	04	05	06	07	08	09	
S	1	2	2	2	5	0	2	2	7	23
N	2	4	8	4	12	6	8	4	4	52
R	1	1	2	2	2	2	2	3	1	16
Total	4	7	12	8	19	8	12	9	12	91

Hypothetical PSU Selection



Hypothetical PSU Selection (continued)

	CBSA	2010 Population	Probability of Selection
✓	Augusta, GA-SC	564,873	0.9221
	Jessup, GA	30,099	0.0491
	Fitzgerald, GA	17,634	0.0288
	Total	**612,606**	**1.0000**

	CBSA	2010 Population	Probability of Selection
	Columbus, GA-AL	294,865	0.4783
	Valdosta, GA	139,588	0.2264
✓	LaGrange, GA	67,044	0.1088
	Moultrie, GA	45,498	0.0738
	Douglas, GA	42,356	0.0687
	Thomaston, GA	27,153	0.0440
	Total	**616,504**	**1.0000**

Number of Addresses

- **Local Target Sample Size**
  - Allocate 12,000 addresses in each survey to individual PSUs, proportional to each stratum's population
  - Minimizes CE's nationwide variance

Number of Addresses (continued)

Given the values of p_i and r_i for every index area i, find the values of n_i that	
Minimize	$\sum_{i=1}^{91} \left(\frac{n_i r_i}{NR} - \frac{p_i}{p} \right)^2$
Subject to:	$\sum_{i=1}^{91} n_i = 12{,}000$
	$n_i \geq 0$, for $i = 1$ *to* 91

Number of Addresses (continued)

where

- p_i = population of the *i*-th index area;
- r_i = productivity rate (eligibility rate times the response rate) of the *i*-th index area;
- n_i = number of addresses allocated to *i*-th index area;
- p = $\sum_{i=1}^{41} p_i$ is the population of the United States;
- $n_i r_i$ = expected number of interviewed households in the i-th index area;
- NR = $\sum_{i \in USA} n_i r_i$ is the expected number of interviewed households nationwide.

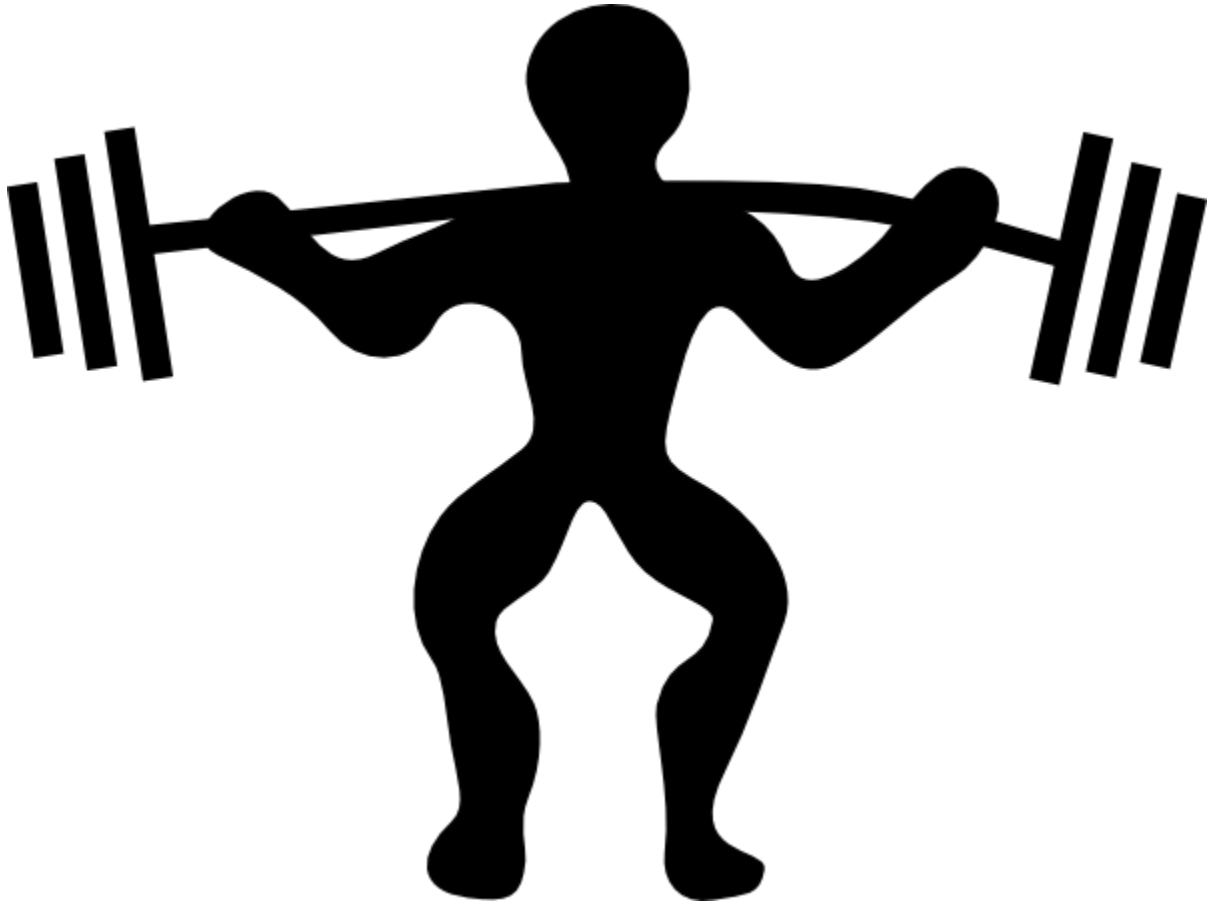
Calculating the Productivity Rate

- 83% **Eligibility** rate
  - (most of the missing 17% are unoccupied)
- 60% **Response** rate
- 50% **Productivity** rate ($0.50 \approx 0.83 \times 0.60$)

Selecting a Random Sample of Households

- Sort households from poor to rich based on information from Decennial Census and ACS
- Compute the sampling interval for each PSU
  - Sampling interval = (# addresses in sampling frame) ÷ (# addresses in CE sample)
- Typical sampling intervals:
  - Every 1,000th address (N and R PSUs)
  - Every 5,000th address (S PSUs)

Weighting Process



Weighting Process

- Base Weight (~10,000)
  - Household + 9,999 others
- Non-interview Adjustment Factor (~1.75)
  - Type A: Refusal to Participate
- Calibration Adjustment Factor
  - Adjusts sample estimate to CPS Totals
  - About 1.15 for Interview Survey

Weighting Process: Calculating the Base Weight

(using hypothetical values)

- PSU Population 538,200
  - MAF counts 224,250 housing units
  - 115 addresses allocated for each survey
  - "Take Every" = 224,250 / 115 ≈ 1,950
- Stratum population 2,800,000
- PSU Weight = 2,800,000 / 538,200 ≈ 5.2025
- Base Weight = "Take Every" * PSU Weight

  ≈ 1,950 * 5.2025 = 10,145

Weighting Process: Calculating the Final Weight

- Variable FINLWT21

- = Base Weight
  - x Non-Interview Adjustment Factor
  - x Calibration Adjustment Factor

- Around 20,000 for Interview Survey, 40,000 for Diary Survey

Conclusion

- Both Sample Design and Weighting Work Together to Produce:
  - Unbiased Estimates of U.S. Expenditures
  - Subject to Allotted CE Budget

Contact Information

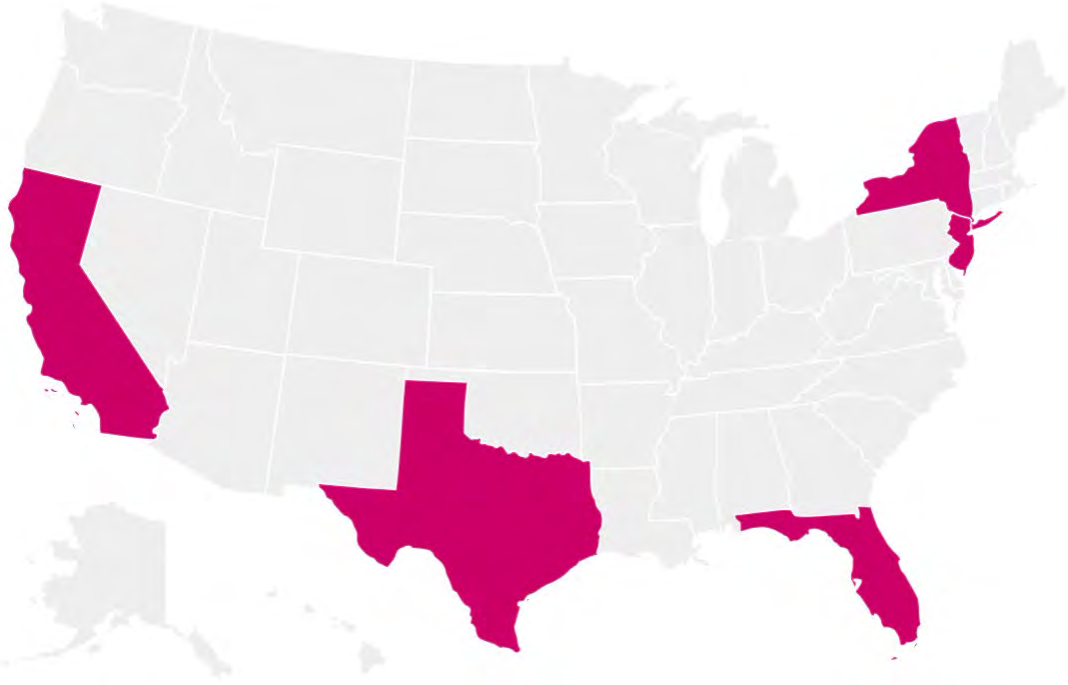
Brian T. Nix
Mathematical Statistician
Statistical Methods Division
www.bls.gov/cex
202-691-6877
Nix.Brian@bls.gov

Consumer Expenditure Surveys: New Geographic Data

Taylor J. Wilson
Economist
Bureau of Labor Statistics
Microdata Users' Workshop
18 July 2019

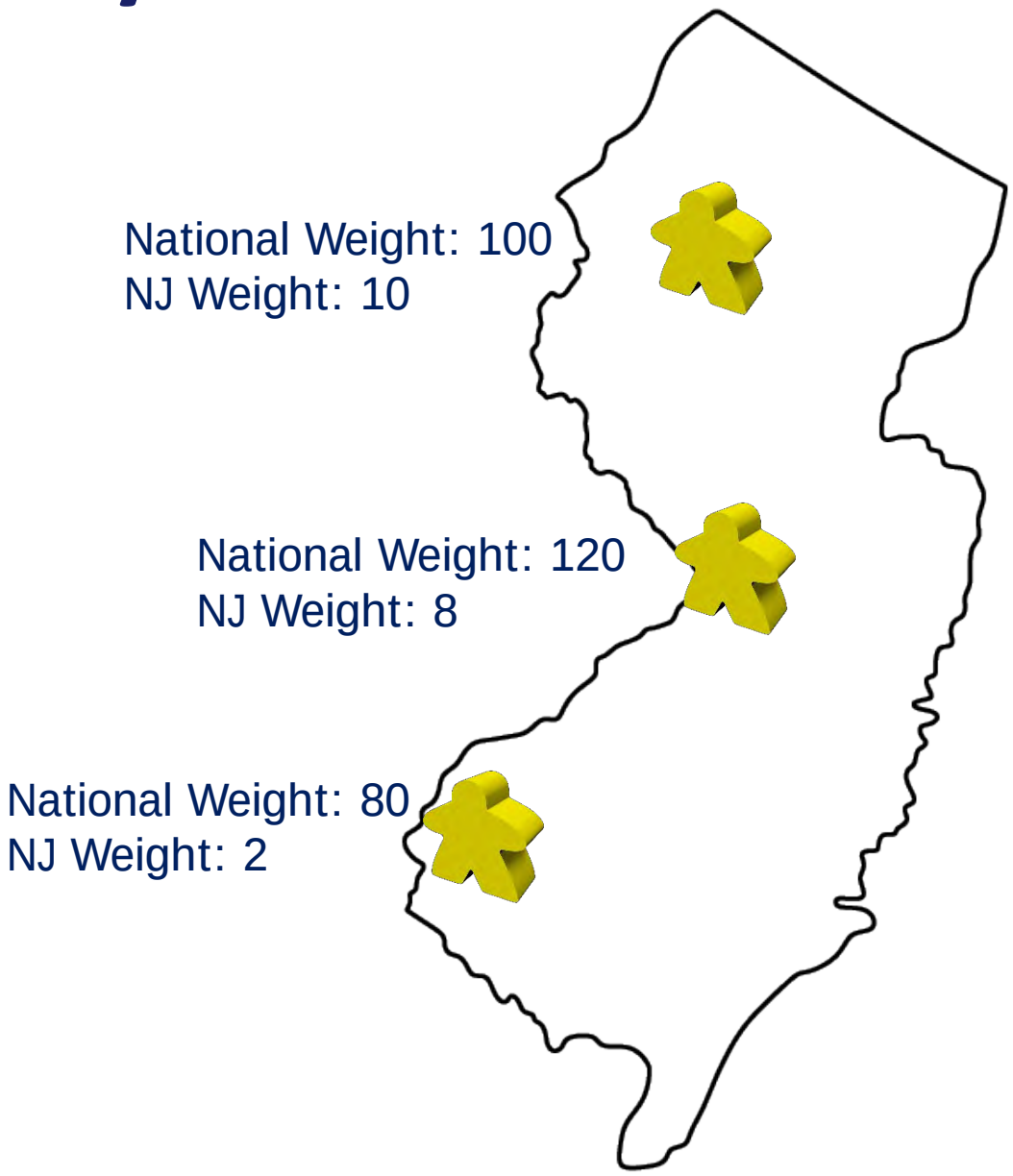
State Estimates

- First test estimates produced for three states using 2013 data.
- First official weights published using 2016 data for three states in 2018.
- Two additional states were added in 2019 using the 2017 data.



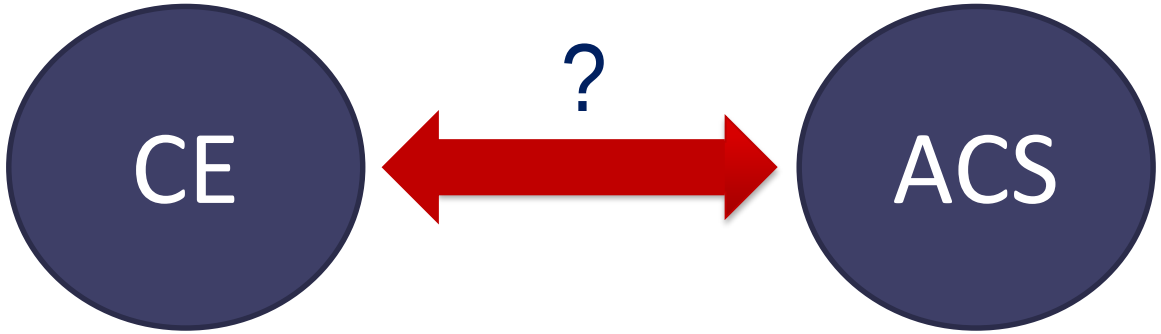
How do they work?

- Suppose the population of New Jersey was 20.
- Each consumer unit in New Jersey is re-weighted to represent the population of New Jersey *instead of* the national population.

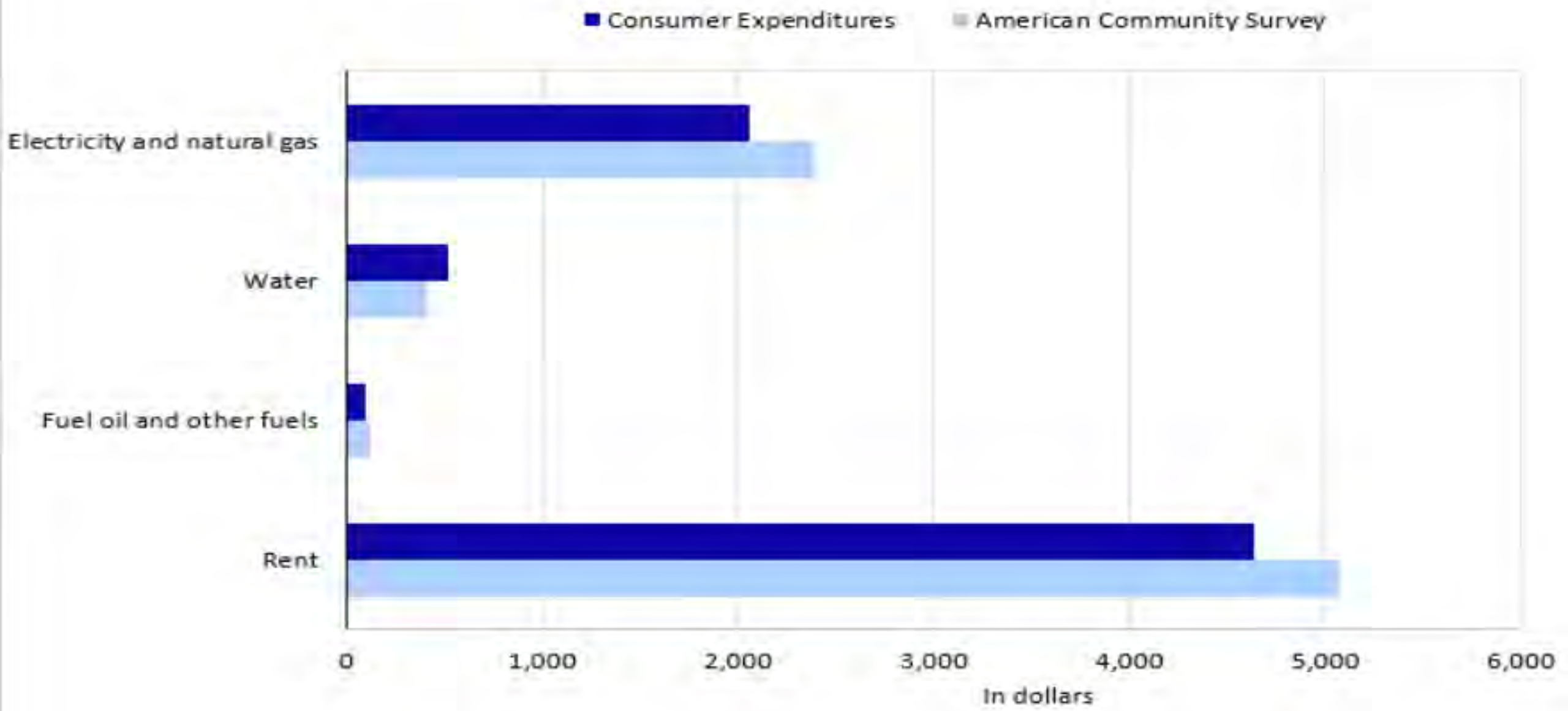


How well do they work?

- Use of other sources which already have state weights for comparison. (ACS)
- Does the state weight produce a difference from other sources that is equal to or better than the national weight?

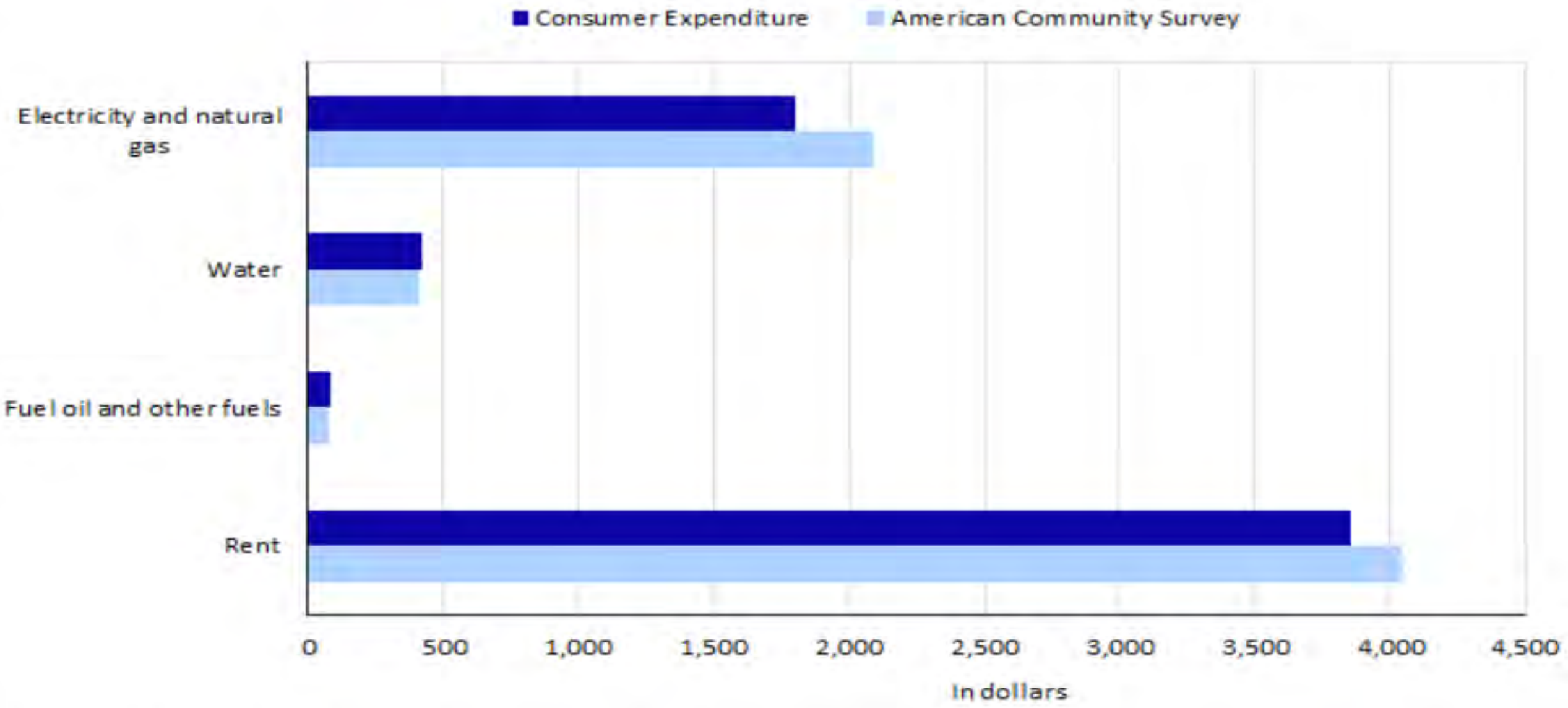


CE-ACS Comparison using New Jersey Weights from Each Survey, 2016



Source: U.S. Bureau of Labor Statistics, Consumer Expenditure Surveys Public-Use Microdata.

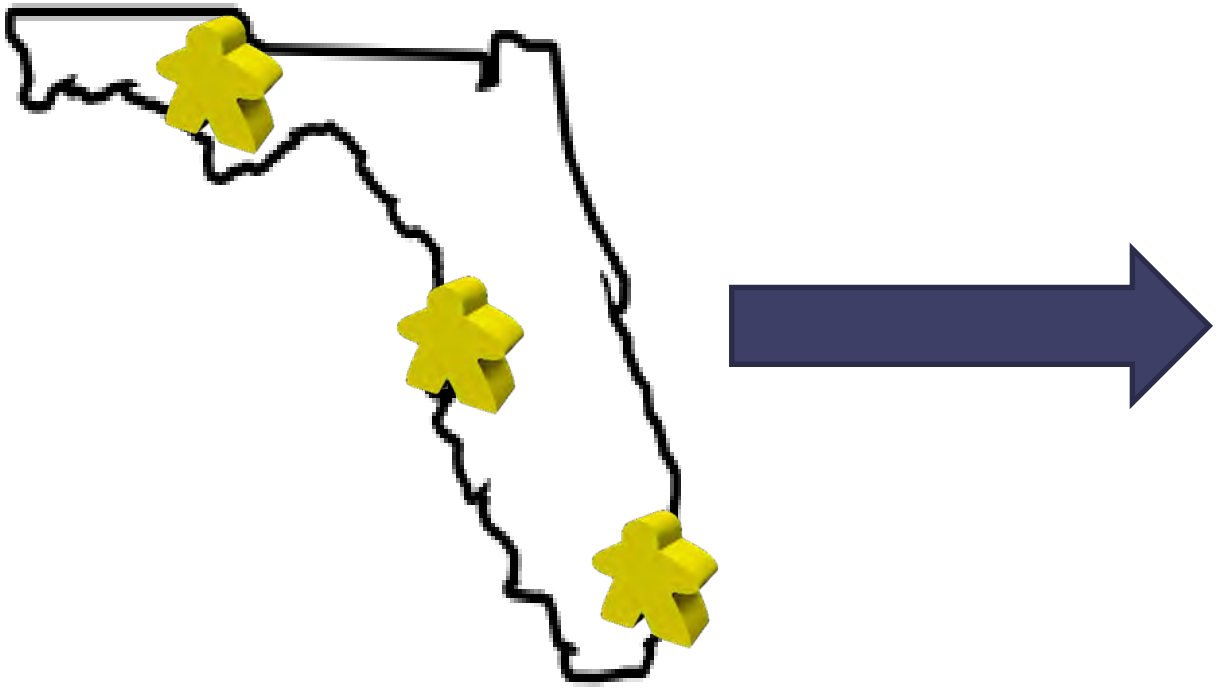
CE-ACS Comparison using National Weights from each Survey, 2016



Source: U.S. Bureau of Labor Statistics, Consumer Expenditure Surveys Public-Use Microdata.

How well do they work?

- Compare population targets to US Census.



1. Add up the weights for Florida from PUMD.
~8.5 Million Consumer Units (CU)

2. Compute average number of persons per CU with the weight.
2.4 persons per CU

3. Multiply and compare to census value.
8.5 * 2.4 = 20.4 million persons

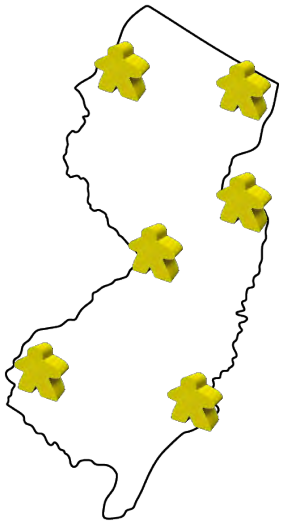
Approximately equal to the 20.66 reported Florida population.

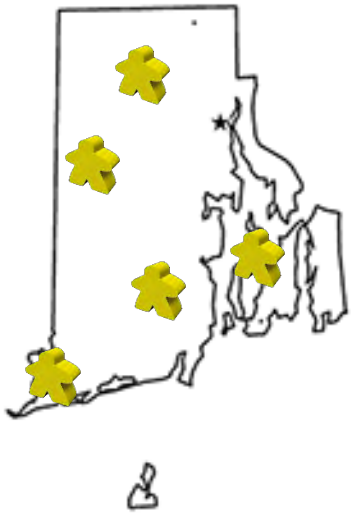
Who uses them?

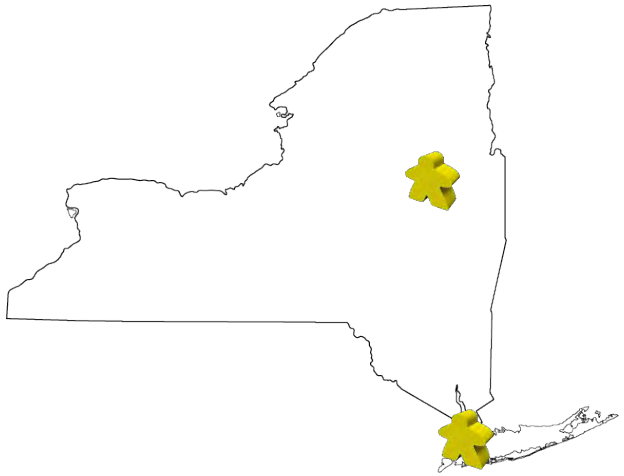
- Academics
  - Researchers have been asking for greater geographic detail for a long time. This is an effort to provide it where we can.
- Government
  - The New Jersey weights were utilized by the New Jersey State Government.
- You!
  - Curious individuals can now answer questions about state level expenditures.

Do you have an example?

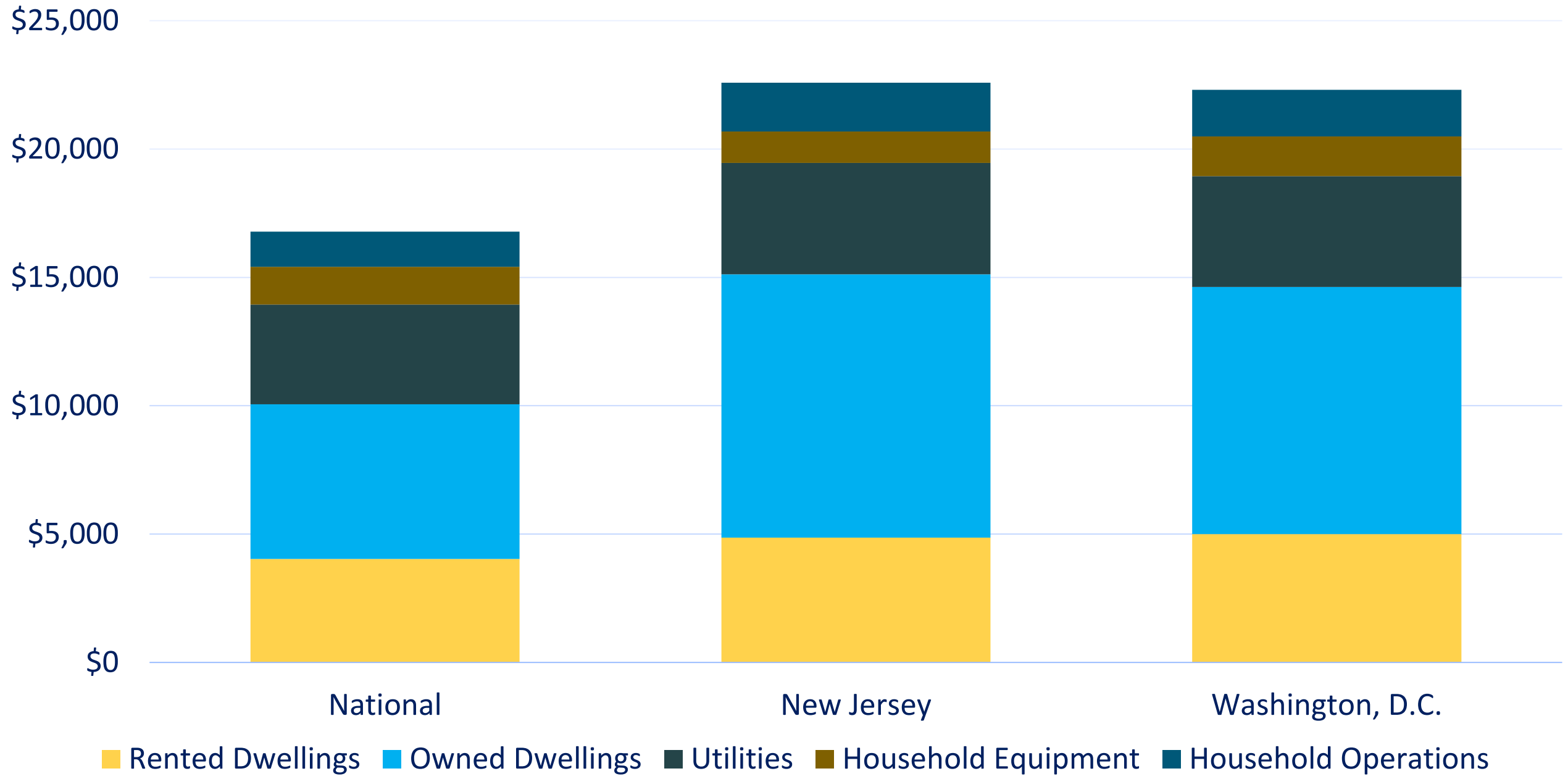
- Use case to do a static examination of the housing market in New Jersey.
- New Jersey has the highest population density of any state.



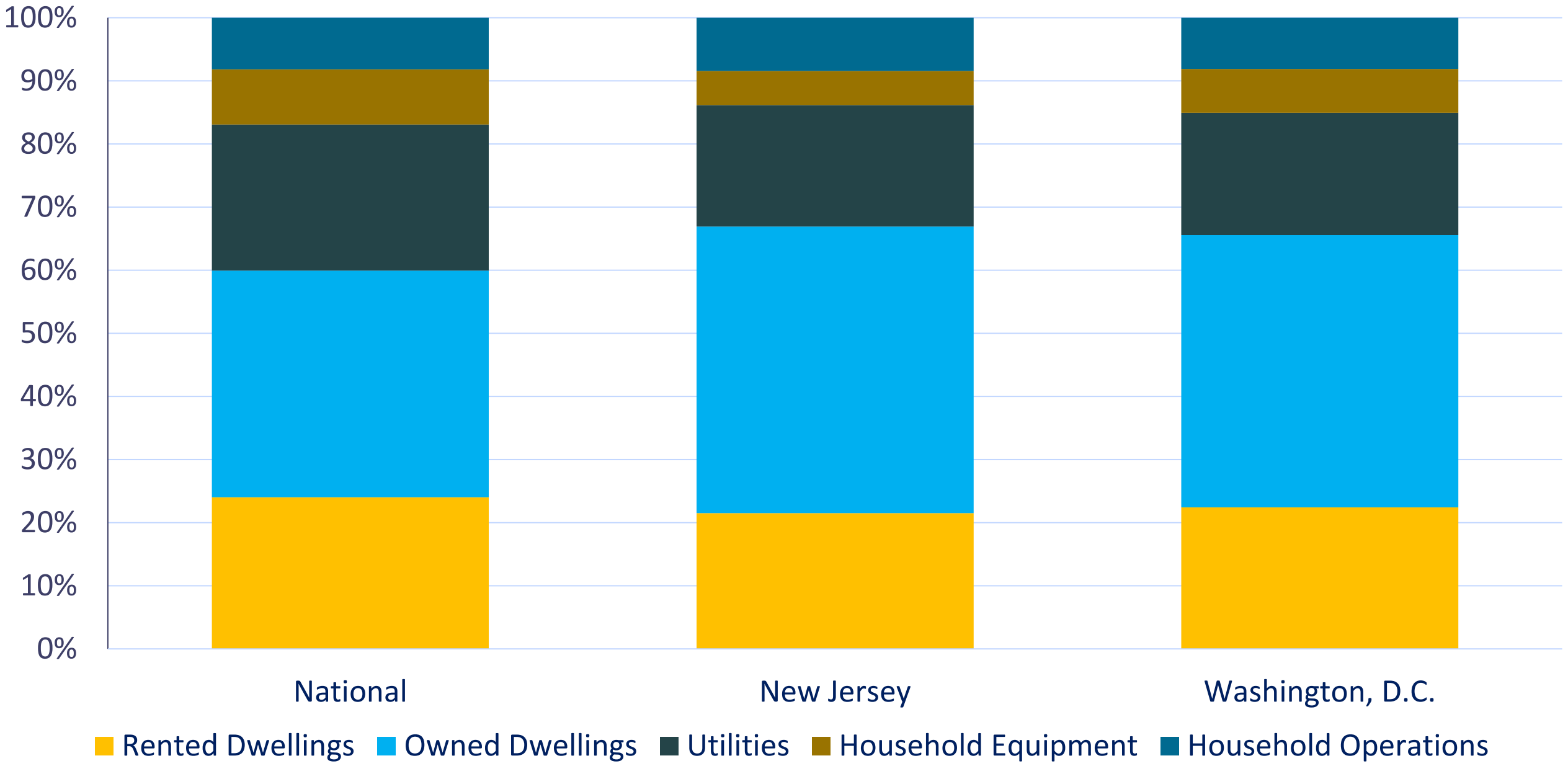




Housing Expenditures and Components by Selected Geographies

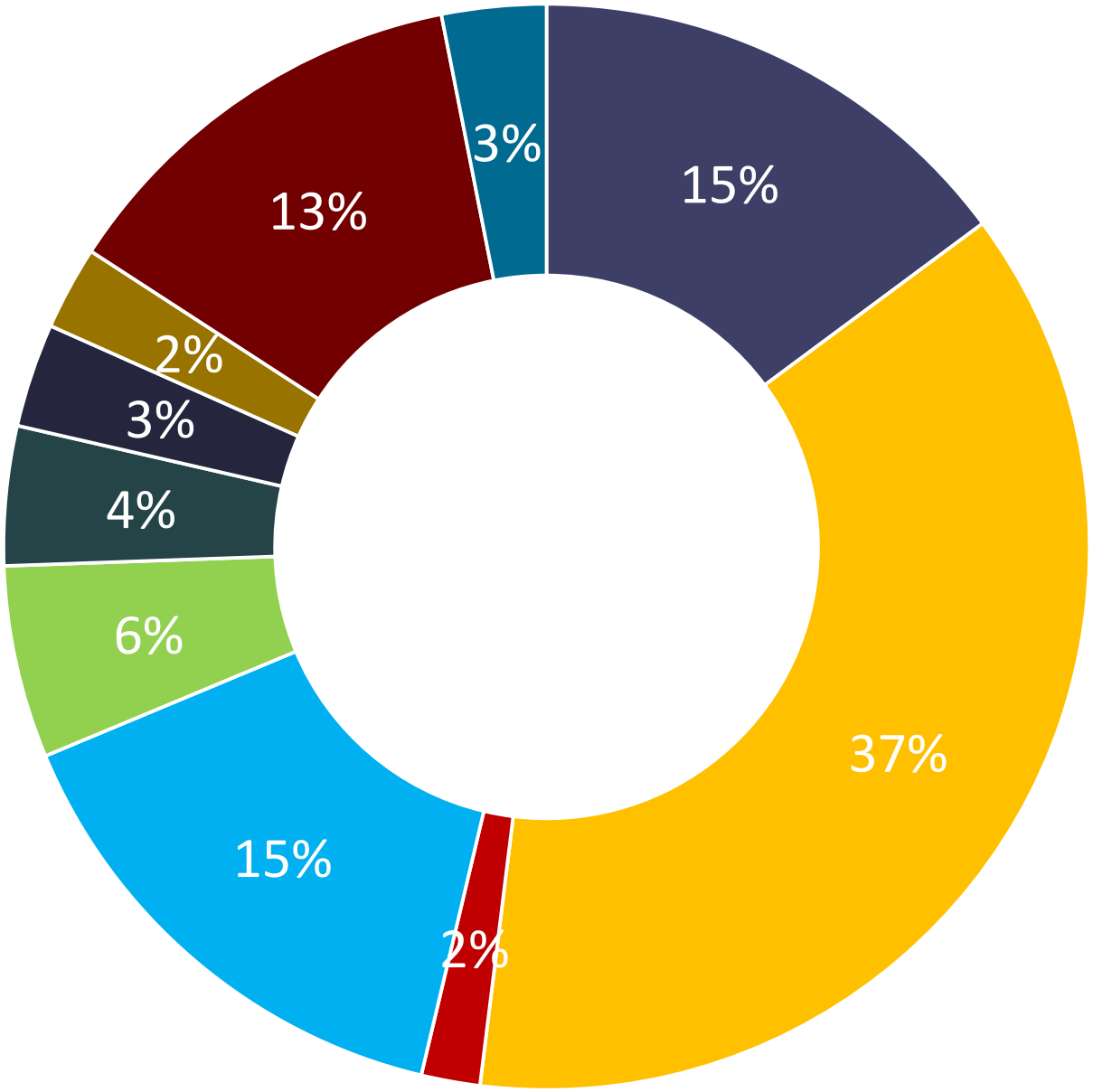


Proportion of Housing Expenditures and Components by Selected Geographies



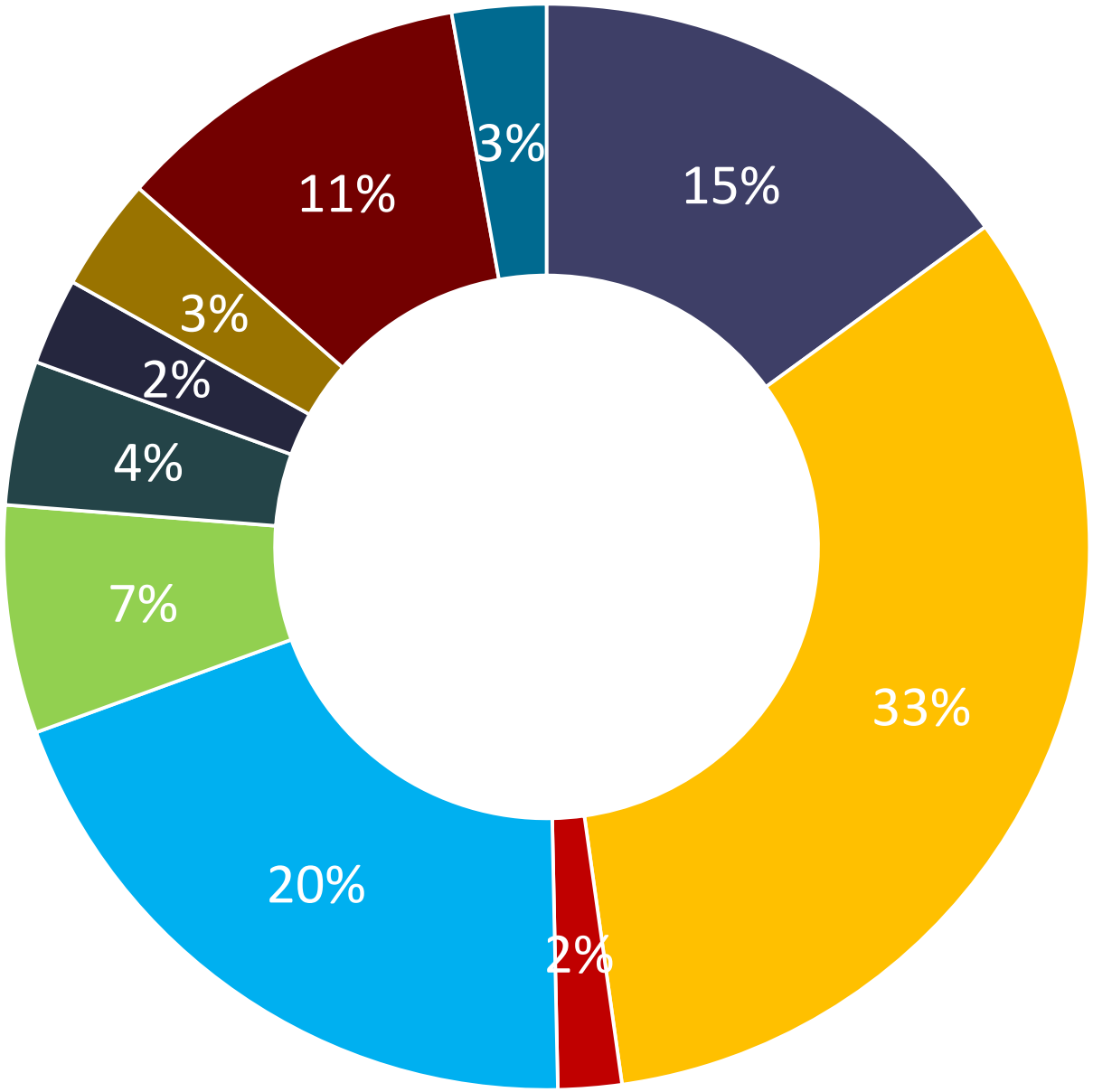
Other Uses?

- Creating state-level market baskets
- Has the potential to calculate cost weights for future consumer price indexes.
- Can be used to evaluate more local state-level decisions



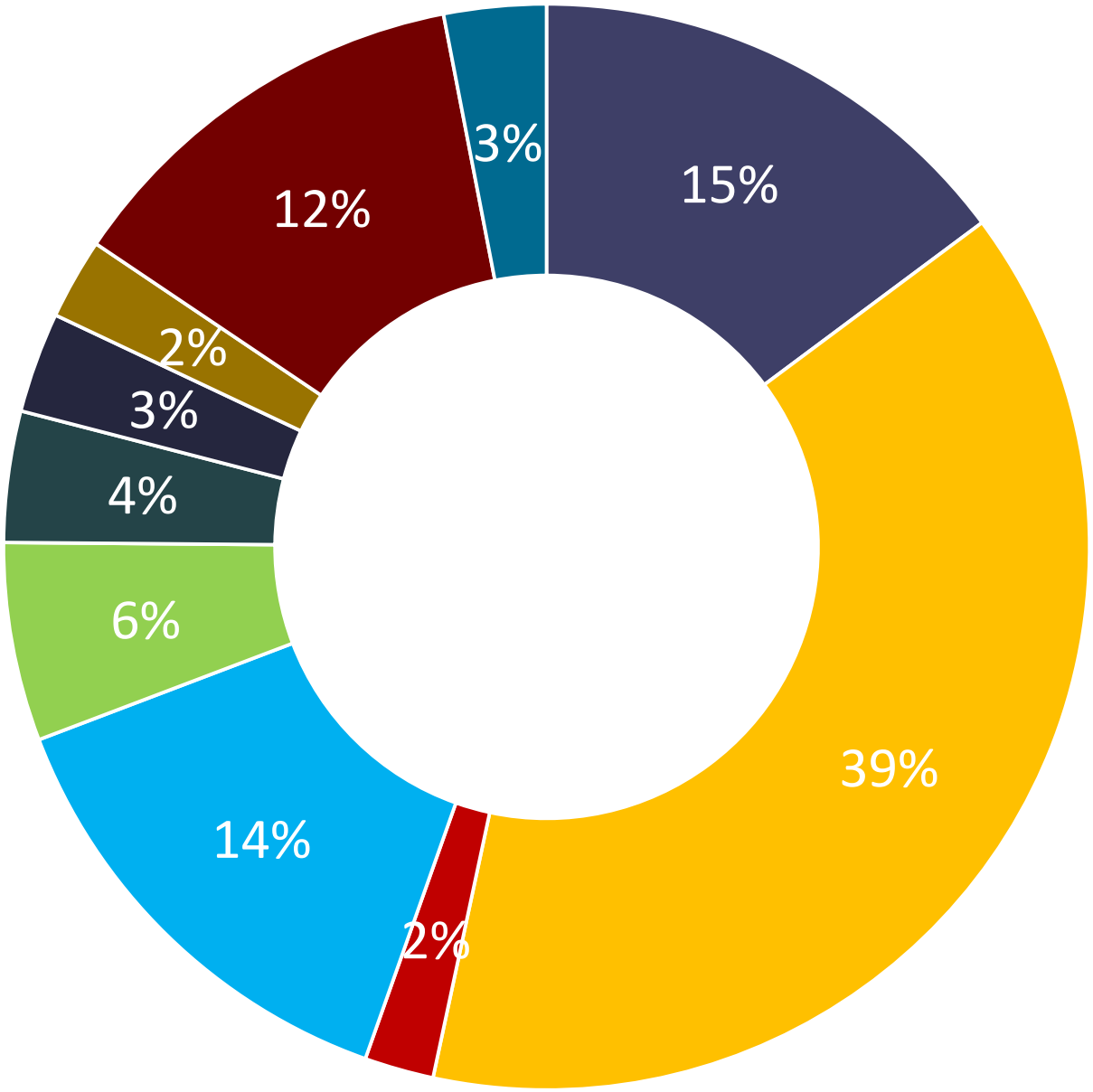
California, 2017

- Food
- Housing
- Apparel and services
- Transportation
- Healthcare
- Entertainment
- Education
- Cash contributions
- Personal insurance and pensions
- All other expenditures



Texas, 2017

- Food
- Housing
- Apparel and services
- Transportation
- Healthcare
- Entertainment
- Education
- Cash contributions
- Personal insurance and pensions
- All other expenditures



New York, 2017

- Food
- Housing
- Apparel and services
- Transportation
- Healthcare
- Entertainment
- Education
- Cash contributions
- Personal insurance and pensions
- All other expenditures

Shares of Total Expenditure, by Major CPI Item Categories
2017 data, Texas

	Texas Shares	National Shares	Absolute Difference
Food and Beverages	0.15	0.13	0.02
Housing	0.33	0.33	0.00
Apparel	0.02	0.03	0.01
Transportation	0.20	0.16	0.04
Medical Care	0.07	0.08	0.01
Recreation	0.04	0.05	0.01
Education	0.03	0.02	0.01
Other Goods and Services	0.16	0.20	0.04

What's next?

- Every state is being evaluated for its potential to generate a weight.
- The following concerns are evaluated,
  - Sample size
  - Confidentiality
  - Long term retention in the survey

Contact Information

Taylor J. Wilson
(202) 691-6550
Wilson.Taylor@bls.gov

Comparing Selected Expenditures of Dual and Single Income Households with Children

Julie Sullivan
Economist
Division of Consumer Expenditure Surveys
Microdata Users' Workshop
July 18, 2019

Outline

1. Motivation
2. Sample Groups
3. Selected Expenditures
4. Preliminary Results

Outline

1. Motivation
2. Sample Groups
3. Selected Expenditures
4. Preliminary Results

Dual Income Families as Majority

- CE data show the majority of households in America are dual income
- Shift from single to dual income as the majority in 1980s (Pew Research Center)
- Would like to compare the different groups' spending habits

Outline

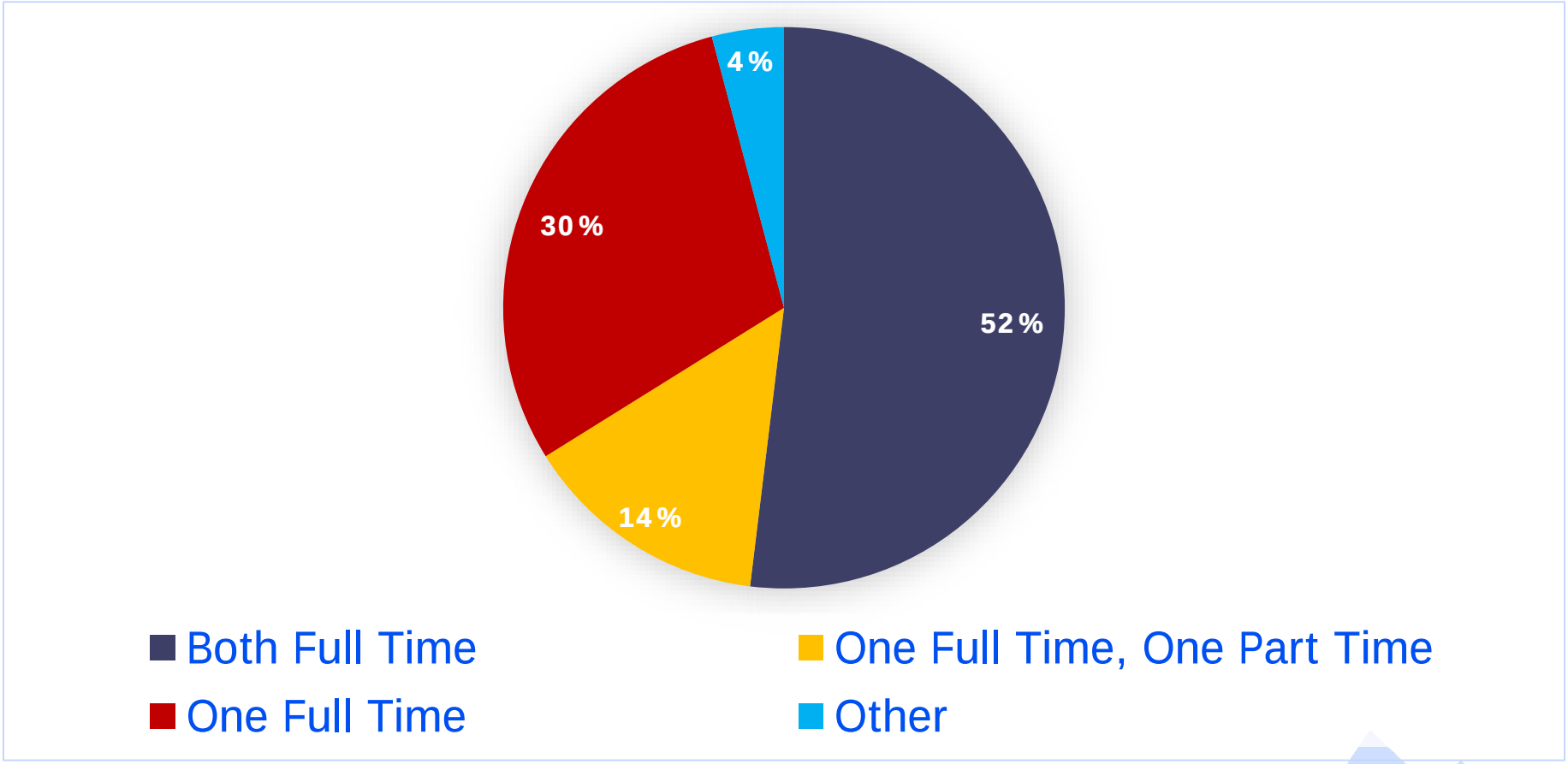
1. Motivation
2. Sample Groups
3. Selected Expenditures
4. Preliminary Results

Subgroups

Two parent households (2015-2017):

- Employment type
  - Whether one or both spouses worked full time during the previous year
  - Also if one spouse worked full time and the other part time during the previous year
- Child age
  - Describes the range in which all children in a household fall
  - Controls for differences in families' expenditures based a children's ages

Percentage by employment type, 2015-2017



Number of Households, Unweighted

	One Full Time	One Full Time, One Part Time	Both Full Time	Other	Total
Oldest child < 6	771	274	1,047	73	2,165
All children age 6-11	340	201	665	30	1,236
All children age 12-17	435	258	914	110	1,717
Total	1,546	733	2,626	213	5,118

Number of Households, Weighted (in millions)

	One Full Time	One Full Time, One Part Time	Both Full Time	Other	Total
Oldest child < 6	15.9	5.8	21.8	1.5	45.0
All children age 6-11	6.6	4.1	13.9	0.6	25.2
All children age 12-17	8.9	5.2	19.2	2.3	35.6
Total	31.4	15.0	54.9	4.4	105.7

Percentage by employment type and child age

	One Full Time	One Full Time, One Part Time	Both Full Time	Other
Oldest child < 6	35.3%	12.8%	48.4%	3.4%
All children age 6-11	26.4%	16.2%	55.1%	2.3%
All children age 12-17	25.0%	14.5%	54.1%	6.4%

Mean number of children by employment type and child age

	One Full Time	One Full Time, One Part Time	Both Full Time
Oldest child < 6	1.59	1.55	1.39
All children age 6-11	1.67	1.64	1.49
All children age 12-17	1.47	1.48	1.44

Outline

1. Motivation
2. Sample Groups
3. Selected Expenditures
4. Preliminary Results

Analyzed Expenditures

Diary:

- Food At Home (groceries)
- Food Away From Home (dining out)

Interview:

- Transportation
- Education
- Childcare

Food

- Hypothesis: Dual income families spend more on food away from home on average than single income families
  - Trade off between time and money
  - American Time Use Survey (ATUS)
  - Perhaps smaller budget share goes to food away from home compared to single income families

Transportation

- Hypothesis: Dual Income families spend more on transportation on average because of commuting
  - Gas
  - Public Transportation Fares
- Perhaps offset by single income families running errands

Education

- Hypothesis: Dual Income families spend more on private school education
  - Private tuition for elementary and high school

Childcare

- Question: How do childcare expenses compare to a potentially higher income?
  - Compare total income
  - Use outlays for a proxy for permanent income

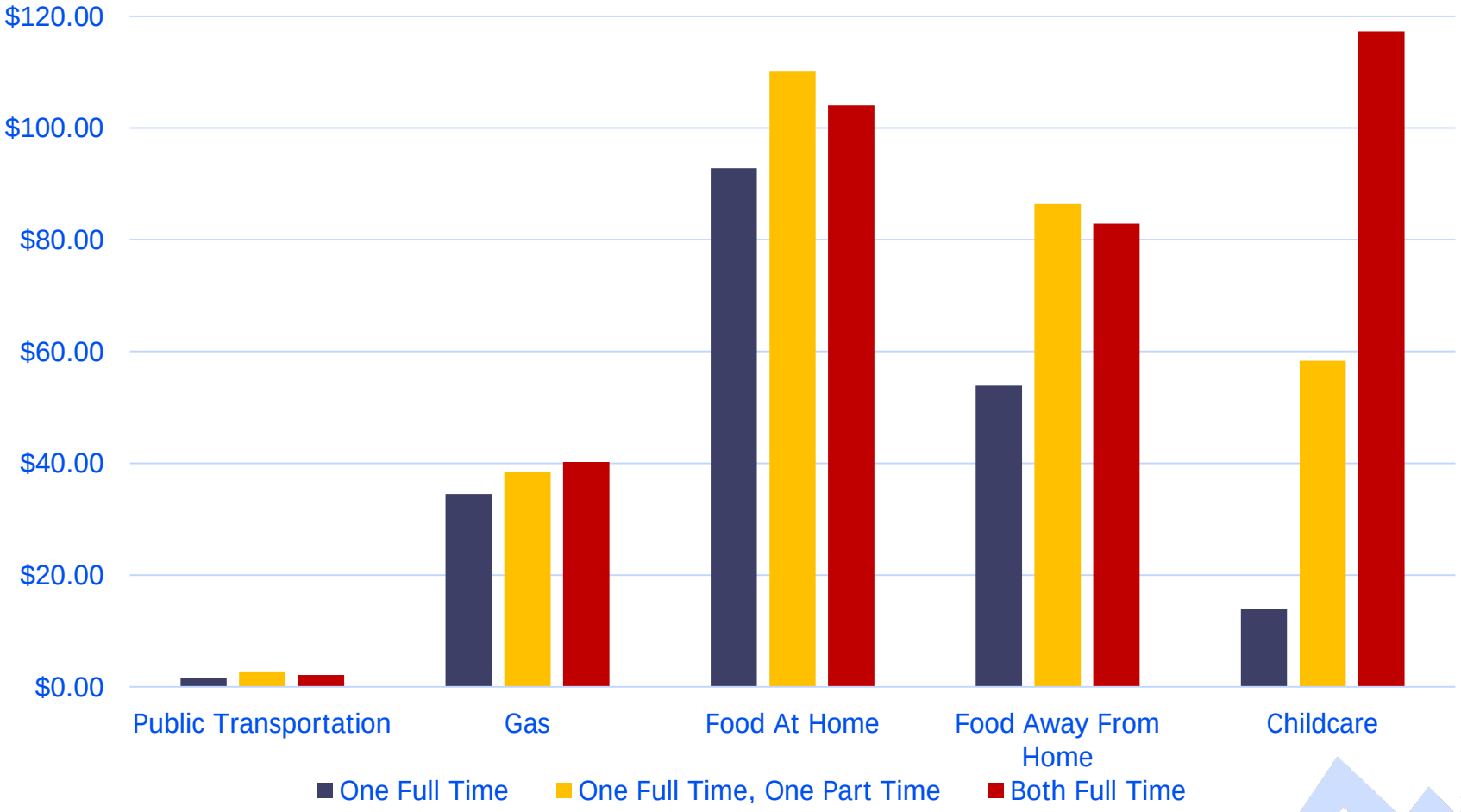
Outline

1. Motivation
2. Sample Groups
3. Selected Expenditures
4. Preliminary Results

Mean Weekly Expenditures of Families with All Children Younger than 6

	One Full Time	One Full Time, One Part Time	Both Full Time
Public Transportation	$1.52	$2.64	$2.12
Gas	$34.52	$38.47	$40.22
Food At Home	$92.78	$110.20	$104.04
Food Away From Home	$53.89	$86.36	$82.89
Childcare	$13.96	$58.35	$117.28

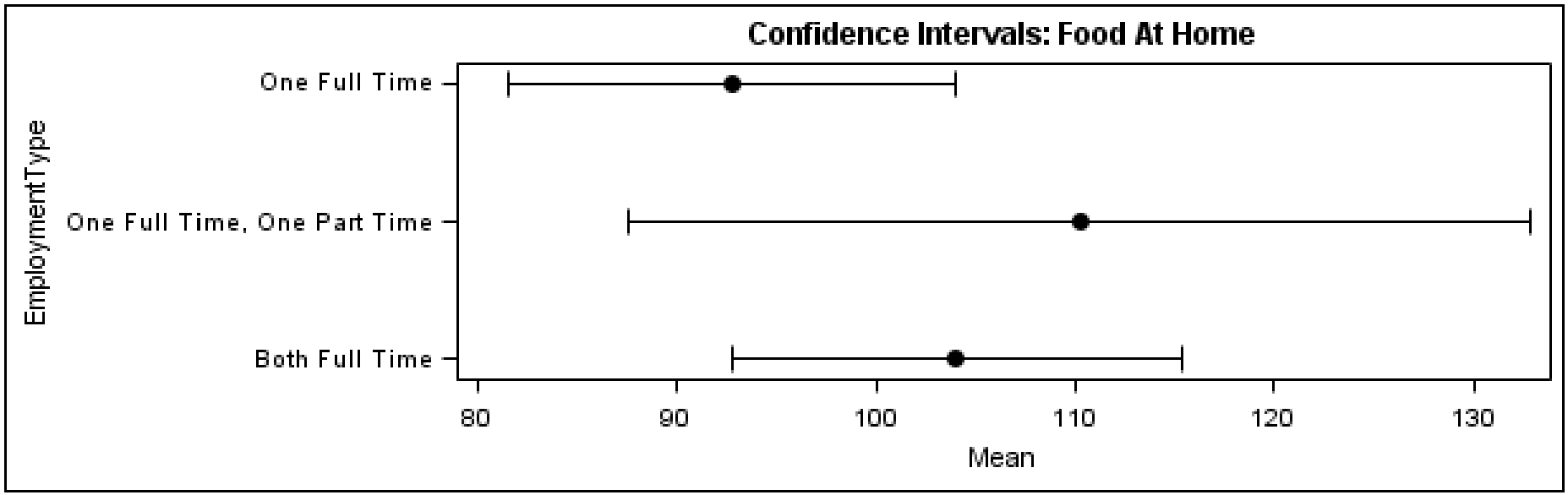
Graph of Selected Mean Weekly Expenditures



Differences in Food at Home and the Statistical Significance

		Mean Difference	Std Error	t Value	Pr > |t|
Both Full Time	One Full Time, One Part Time	-6.15	13.85	-0.44	0.659
Both Full Time	One Full Time	11.27	7.75	1.45	0.153
One Full Time, One Part Time	One Full Time	17.42	12.37	1.41	0.166

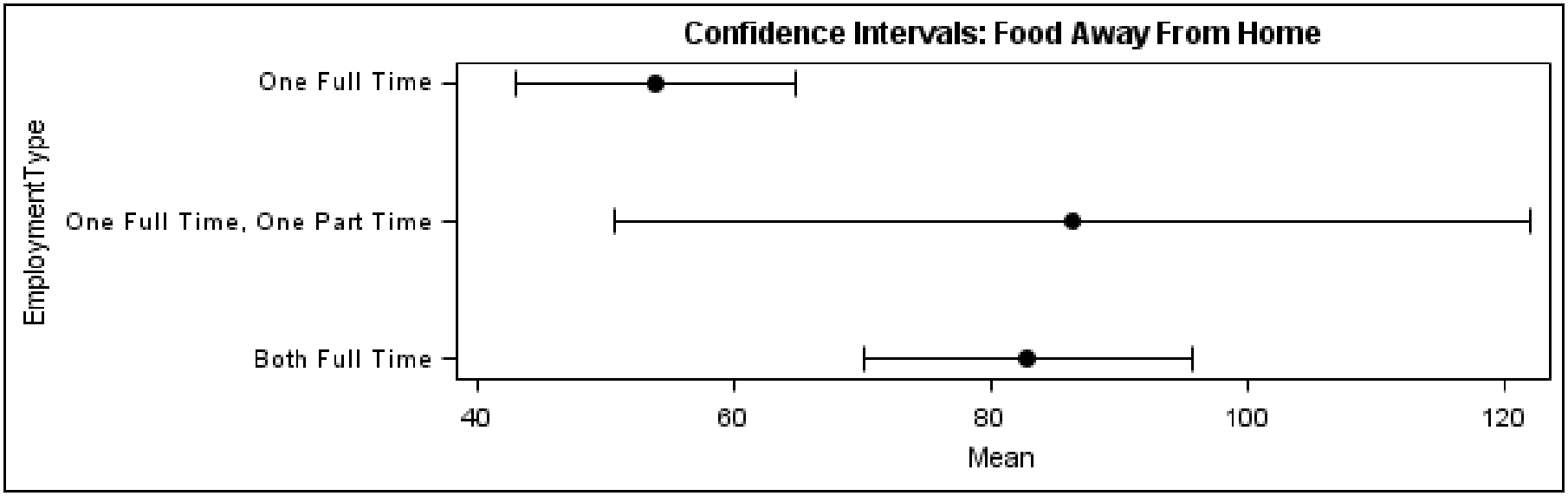
Confidence Intervals: Food At Home



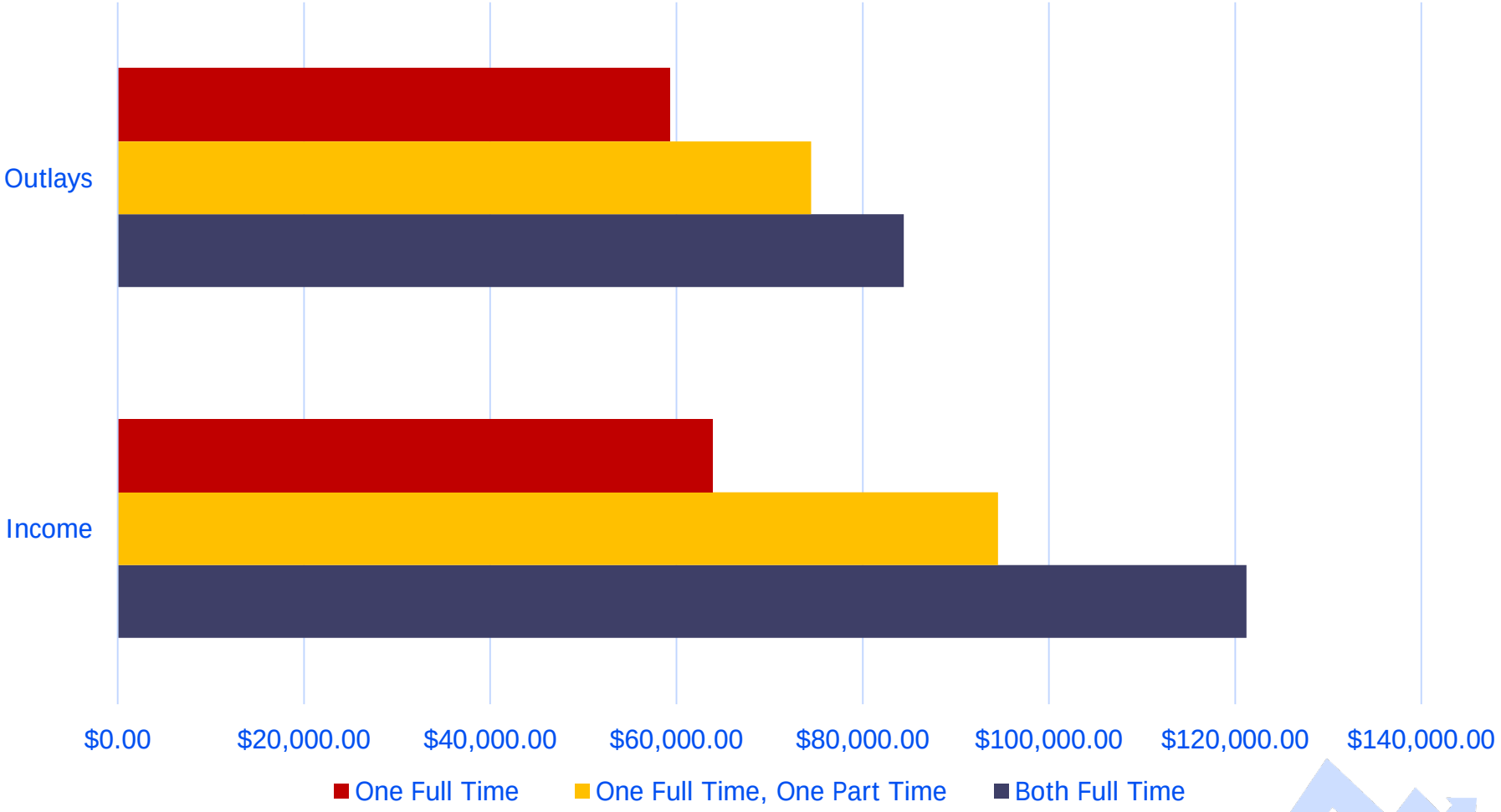
Differences in Food Away From Home and the Statistical Significance

		Mean Difference	Std Error	t Value	Pr > |t|
Both Full Time	One Full Time, One Part Time	-3.47	19.22	-0.18	0.858
Both Full Time	One Full Time	29.01	9.03	3.21	0.003
One Full Time, One Part Time	One Full Time	32.48	18.70	1.74	0.090

Confidence Intervals: Food Away From Home



Mean Outlays and Income of Families with Children all younger than 6



Summary of Preliminary Results: Expenditures that are statistically significantly different

	One Full Time	One Full Time, One Part Time	Both Full Time
Gas	$34.52	$38.47	$40.22
Food Away From Home	$53.89	$86.36	$82.89
Childcare	$13.96	$58.35	$117.28

Questions

Contact Information

Julie Sullivan
Economist
Division of Consumer Expenditure Surveys
www.bls.gov/cex
202-691-6234
sullivan.julie@bls.gov

Sales Tax in CE Data

Barbara Johnson-Cox
Economist
Division of Consumer Expenditure Surveys

CE Microdata Users Workshop

July 17 – 19, 2019

Outline

- Sales Tax Rate Edit – Overview
- Sales Tax – Interview Survey
- Sales Tax – Diary Survey
- Sales Tax – Applying Sales Tax Rates
- Sales Tax Rate – Resource Sites

Sales Tax Edit - Overview

Purpose:

- The *basic function* of this edit is to add sales taxes to selected cost fields where the "sales tax" was not included.

- Sales taxes are calculated by applying the sales tax *rate* to the reported *cost* of the item.

Sales tax = Cost * rate

Sales Tax – Interview Survey

- The respondent is *"asked"* **if** the *reported cost* included "sales tax" (for specific items).
  - Household appliances and equipment
  - Household item repairs and service contracts
  - Household furnishings
  - Clothing, clothing services, watches, jewelry
  - Owned vehicles
  - Vehicle maintenance and repair
- If the response is "No" the sales tax edit will add sales taxes to the reported cost *and* change the "No" answer to "Yes" (*sales tax is included*).

Sales Tax – Diary Survey

- Respondents are instructed NOT to include any sales taxes in Parts 1, 3, and 4 of the Diary.
  - Part 1. Food and Drinks for Home Consumption
  - Part 3. Clothing, Shoes, Jewelry, and Accessories
  - Part 4. All Other Products, Services, and Expenses
- The *sales tax edit* will add "sales taxes" to the cost fields of the items that are reported.

Sales Tax – Diary Survey

- In Part 2 of the Diary – respondents are instructed to include the "sales taxes" in the total cost.
  - Part 2. Meals, Snacks, and Drinks Away from Home

- The sales tax edit will "assign" a sales tax "rate" to each item.

Applying Sales Tax Rates

- CE “Sales tax rates” includes a combined total of the State, County and City sales tax rates.

- States can have *multiple* sales tax rates that are applicable to different items.

  State X:

  General Sales Tax Rate = 9%

  Groceries = 3.5%

- The sales tax edit *selects* the “Sales tax rate” by matching the State with the Item – within the Sales Tax Rate files.

Applying Sales Tax Rates

- That "rate" is then applied to the cost to obtain "Total Cost" – which includes the sales tax".

Total Cost = Cost * (1 + RATE)

Ex: $10.60 = $10.00 * (1.06)

- Sales taxes are included in the cost variables in the Microdata.

Both Surveys

- Additional Details:
  - ✓ The total cost is rounded to two decimal places.

Sales Tax Rate Files

- Two files: Interview and Diary
  - Organized by State and Item
  - Updated annually

Sales Tax Rate Resource Sites

- **Sales Tax Rates – Resource Site:**
  - **State Department of Taxation & Revenue**

Contact Information

Barbara Johnson-Cox
Economist
Division of Consumer Expenditure Surveys
www.bls.gov/cex
202-691-6900
cexinfo@bls.gov

Imputation and Allocation of CE Data

Clayton Knappenberger
Senior Economist
Division of Consumer Expenditure Surveys
2019 CE Microdata Users' Workshop
July 19

Outline

1. Process Overview
2. Imputation
3. Allocation
4. Edit Rates and Conclusion

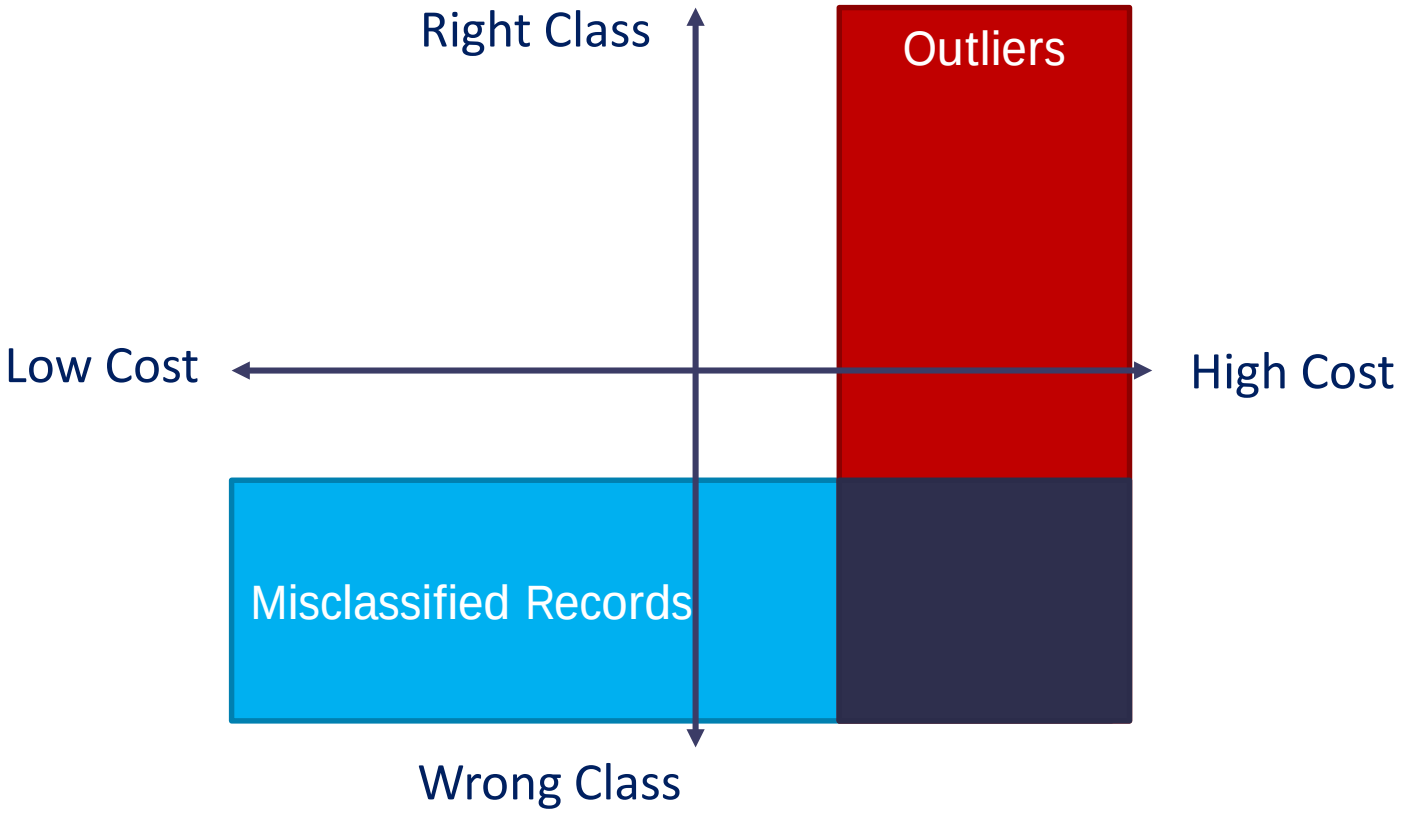
Process Overview

- CE's goal is to map expenditures
  - As monthly amounts
  - To specific Universal Classification Codes (UCCs)
  - In a specific month and year
- However, collected data are often insufficient
  - Collected information has mistakes
  - Respondent does not know or refuses to provide

Process Overview

1. **Data Screening** – check data for errors
2. **Impute** missing values
3. **Allocate** combined expenditures to components for mapping.
4. **Mapping** expenditures to months and UCCs (as well as higher level aggregations)

Data Screening



Misclassifications

- Specific keyword lookups for “hard to classify” items
  - iPad/iPhone/iPod
  - “Glasses”/”Cable”/”Nails”
- General text analysis of item descriptions
- Updates are made based on the reported item description and any interviewer notes

Outlier Review

- Three methods:
  1. Largest Gap – biggest gap between records above the mean
  2. P-Index – value divided by gap minimum
  3. Z-Score – value divided by IQR
- Updates are made by:
  1. Correcting values with available information
  2. Flag the expenditure for imputation

Imputation

1. Weighted Mean Imputation
2. Hot Deck Imputation
   - Use valid records with similar characteristics to replace missing values
3. Percent Distribution Imputation
   - Randomly select a valid value based on the percent distribution of reported values

Weighted Mean Imputation

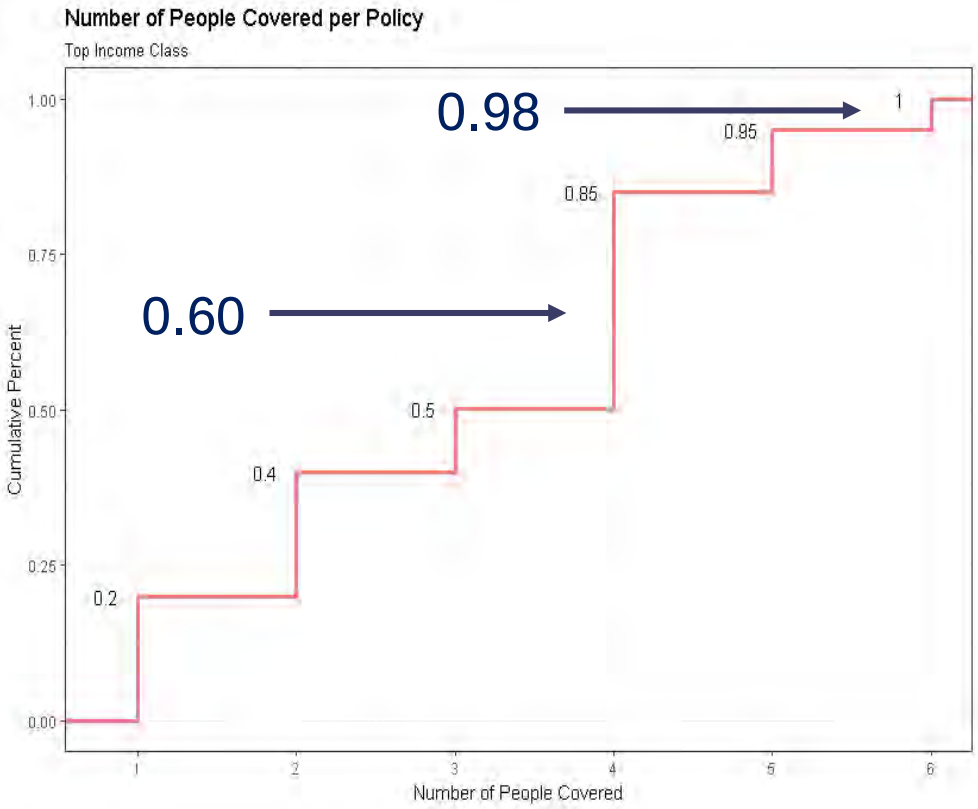
- Use valid records with similar characteristics to define cells
- Calculate the weighted mean of that cell
- Assign the weighted mean of reported expenditures within a given cell to missing or invalid expenditures in the same cell

Hot Deck Imputation Example

- A respondent reports buying a men's jacket, but does not know the cost
- Imputation steps:
  - Select a valid random men's jacket expenditure from all such purchases with the same:
    - Region
    - Area Type
    - Income Class
  - The selected record's expenditure amount is copied to the record being imputed

Percent Distribution

- A respondent does not say how many people are covered by an insurance plan



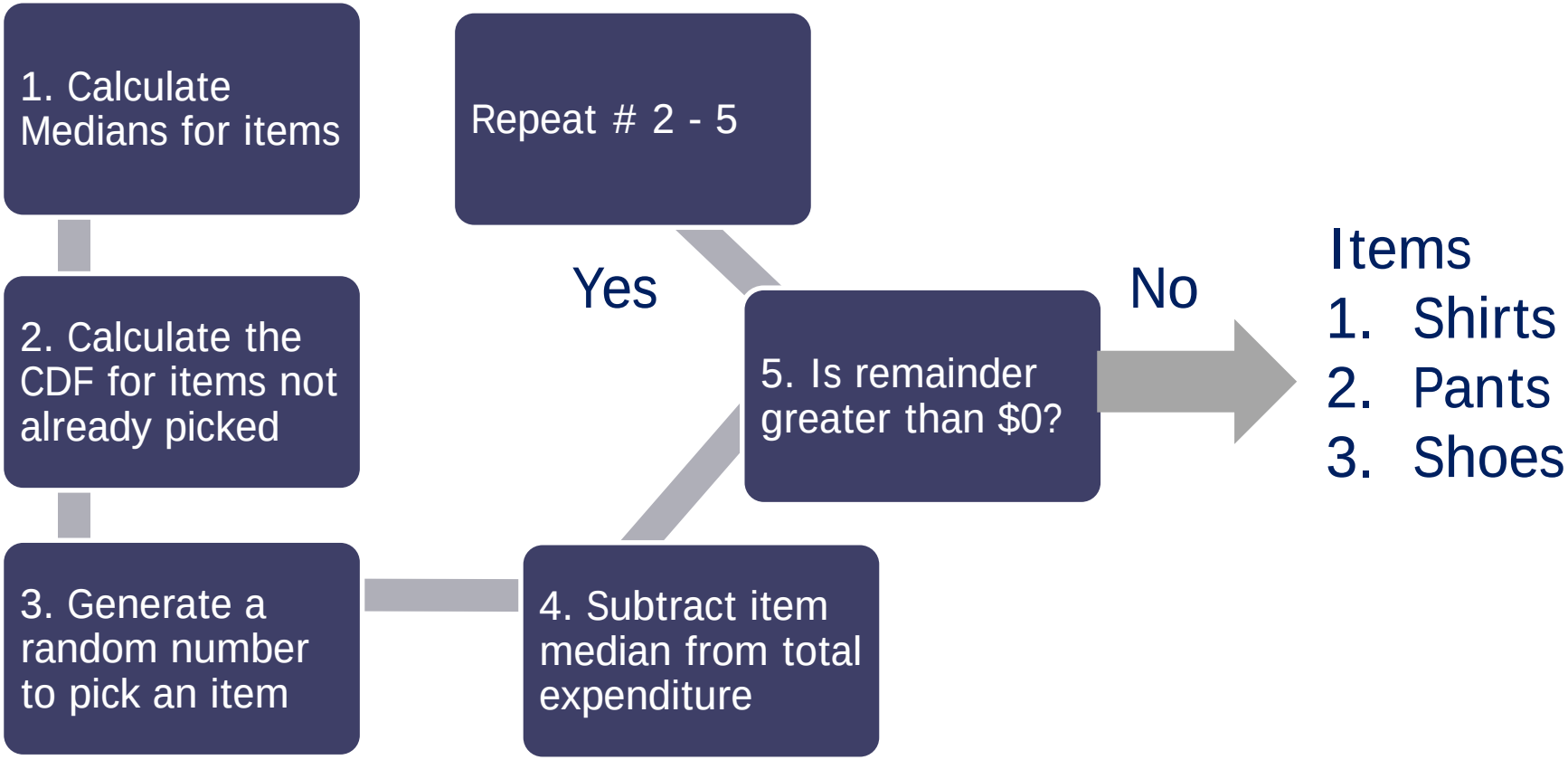
Steps

1. Create CDF
2. Get Random Number
3. Assign value

Allocation

- Example: Respondent reported spending $500 on clothing
- We need two pieces of information:
  1. Targets – shirts, pants, and shoes
  2. Allocation Proportions
     - 45% on shirts
     - 35% on pants
     - 20% on shoes

Picking the Target Items

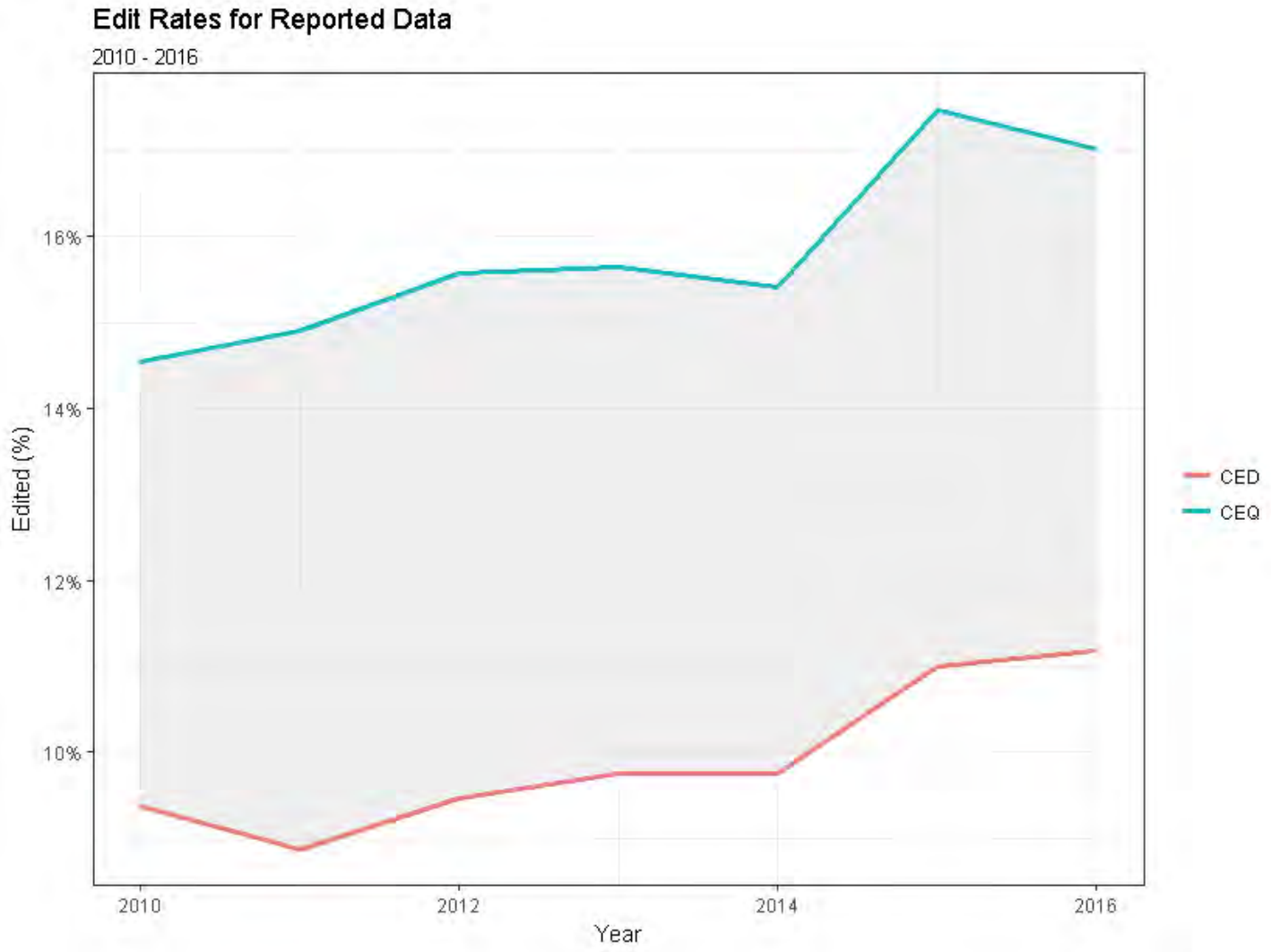


Allocating the amounts

1. Get the weighted means for each item selected
2. Calculate the share of the sum of the means
3. Derive allocation amounts

Item	Mean ($)	Share (%)	Allocation ($)
Shirts	$35.00	21.88%	$109.40
Pants	$67.00	41.87%	$209.35
Shoes	$58.00	36.25%	$181.25
Total	$160.00	100.00%	$500.00

Imputation and Allocation Rates



Why Impute and Allocate?

Benefits

- Meet internal needs for mapping
- Provide complete datasets to users

Concerns

- Our methods rely on MAR assumption
- Potential for underestimated variance

Contact Information

Clayton Knappenberger
Senior Economist
Division of Consumer Expenditure Surveys
www.bls.gov/cex
202-691-6236
knappenberger.clayton@bls.gov

BLS Visiting Researcher Program

Jimmy Choi

Division of Consumer Expenditure Surveys
CE Microdata User's Workshop
July 17-19, 2019

Restricted CE Data

- Restricted CE data are collected, but not publically available
  - Geographic information
    - Census tracks
    - FIPS state and county codes
    - Zip codes
  - Topcoded income and expenditure variables

BLS Visiting Researcher Program: Overview

- Increased interest in using restricted data over the past
- Program allows data users to access restricted data
  - Federal agency
  - Non-profit
  - University or College
- Requires application process

Application Process

1. Contact CE to ensure data are available
2. Explore Visiting Researcher Program Website: http://www.bls.gov/rda/home.htm
3. Submit Application
    - Research Proposal
    - Federal/Non-Federal Visiting Researcher Questionnaire
    - CV/Resume
4. Review by BLS Microdata Access Research Board (MARB)
5. Obtain Title 13 clearance

Time Frame

- Typically takes 3 to 6 months, but can take longer

Jimmy Choi
Economist
Division of Consumer Expenditure Surveys
202-691-7081
Choi.Jimmy@bls.gov

Sneak Peek: Where are we going this year and beyond?

Steve Henderson
Chief, Branch of Information and Analysis
Division of Consumer Expenditure Surveys

July 19, 2019

Celebrating 131 Years
of Consumer Expenditure Surveys!

Vision

The Consumer Expenditure Surveys Program will provide the Nation with accurate and complete information on consumer spending, based on methodologically sound surveys, efficient processing, and timely dissemination

Where are we headed and what have we done lately?

Outline

- What's new for tables and microdata
- Weights for five ~~three~~ selected states
- Next annual release date
- **NEW** Generations and Veteran's Status

Changes for tables and microdata

Recently:

- Redesigned microdata documentation
- **New on September 10**
  -Census Divisions
  -Higher Income cross tabulations

SNEAK PEEK

Coming in 2019 and 2020

- More detailed geographic Census Divisions, in addition to Census Regions, using 2017 and 2018 data
- Pre-1996 Microdata online

Also coming in 2019 and after

- New generational categories, including those born after 1996
- This will split up the Millennials into 1981-1996, and the new Generation Z, born 1997+

- Estimated publication date: September 2020

New Veterans Status

- In addition to asking if anyone in the household is currently in the US military,
- The CE added a new question asking if anyone had served previously
- Started in July 2018 in the Interview survey, and January 2019 for the Diary
- Will be in the ***2019*** PUMD, and in a new table when we have enough data

Going in 2020

- Me, after 42 years with the Federal Government, including 21+ fascinating years with the CE program
- I plan to spend more time with my family, including our first grandchild

Questions?

Contact Information

Steve Henderson
Chief, Branch of Information and Analysis
Division of Consumer Expenditure Surveys
(through the end of December 2019)
202-691-5124
Henderson.Steve@bls.gov

Investigating the Imputation of Assets and Liabilities in the CE Interview Survey

Geoffrey Paulin, Ph.D
Senior Economist
Consumer Expenditure Surveys (CE) Program
CE Microdata Users' Workshop
July 19, 2019
Washington, DC

The Consumer Expenditure Survey (CE) collects information on:

- Expenditures
- Income
- Taxes (income and other)
- Assets and Liabilities

Nonresponse is a problem for each. However, corrections are in place for most of these items:

- Expenditures: Since the 1980s
- Income: Since 2004
- Taxes (income only): Since 2013

Assets and Liabilities are currently under investigation.

Assets and Liabilities

Project Overview

- "The purpose of this team is to initiate and conduct a research project designed to impute missing Interview asset and liability data, leveraging models from income imputation and other relevant procedures."
- "The goal is to implement this into production with 2017 Quarter 2 data."

Source: Charter for the Asset and Liability Imputation Team, 9-9-2014

This presentation describes three aspects of the project:

- What asset and liability data are collected?
- What processes have been considered for imputation of missing values?
- What are the next steps in the investigation?

Asset/Liability Data

Assets:

- Retirement accounts
- Stocks, bonds, mutual funds
- Checking, savings, money market, CDs
- Whole life insurance
- Other, including annuities, trusts, royalties

Liabilities:

- Credit cards
- Student loans
- Other loans, including medical and personal

Collection

- Questions are asked in the final survey (4th interview)
- Most are asked in two parts: Did you have ____? If yes, how much?
- For some items, only a total value is collected. In these cases, it is not clear whether $0 means:
  - No, I did not have such an account or
  - Yes, I had an account, but it is empty.

Collection, continued:

- For each asset/liability, the total value/balance/amount owed is collected:
  - As of today
  - As of one year ago today
- Bracket questions are asked when the respondent cannot provide a specific value.

The team considered several methods:

- Survey of Consumer Finance method (multiple imputation, iterative process)
- Regression trees
- Hotdeck

…But none is feasible.

Going back to the original motivations (charter):

A system based on income imputation processing is being investigated.

- Regression-based, multiple imputation of each component asset/liability, from which "total change in" values can be derived.
- For each component, separate models are run for demographic groups across which large variation in parameter estimates is observed or expected.

For example:

Consider IRAX.

- Amount reported when asked: "What is the total value of all retirement accounts such as 401(k)s, IRAs, and Thrift Savings Plans that you or your household own/owns?"
- Expected to vary considerably by age
- Preliminary tests support use of one model for each age group (group 1: age<A; group 2: A<=age<B, etc.)

Bracket imputation will also be used:

- Respondent identifies range in which asset/liability falls (e.g., less than $X; $X to $Y; etc.)
- Five values are selected based on current methods used in income imputation; each falls within the specified bracket range.
- Open-ended brackets ($Z or over) also are treated in income imputation

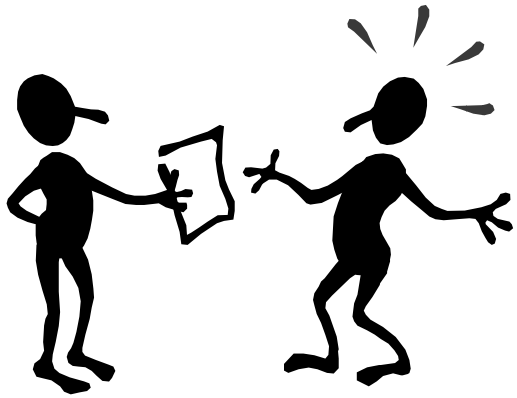
Related Challenges:

- How to distinguish $0 meaning no balance from $0 meaning no account.
- How to identify groups upon which to base models. That is:
  - Grouping variable: Is IRAX (e.g.) based on age, occupation, or something else?
  - Variable range: If age, where do the breaks occur—under 35, 35 to 64, 65 and older, or under 25, 25 to 34, etc.?

Work in progress:

- Identifying groups, and selecting variables to include within each model.
  - ANOVA/Chow tests have been used so far to test differences/pooling potential across groups.
  - Variables used in income imputation are considered the "starter group," with some to be added, deleted, or redesigned. (Example: Age ranges used in binary variables could be widened or narrowed.)

Comments/Suggestions are welcome!



Contact Information

Geoffrey Paulin, Ph.D.
Senior Economist
Consumer Expenditure Survey Program
www.bls.gov/cex
202-691-5132
paulin.geoffrey@bls.gov