

Consumer Expenditure Surveys Public Use Microdata (PUMD) Overview

Bryan Rigg

**Division of Consumer Expenditure Surveys
CE Microdata Users' Workshop
July 18-20, 2023**



What Are Microdata?

- “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.



Consumer Expenditure Surveys

 [Go](#)[CE Home](#)[CE Publications](#)[CE Data](#)[CE Methods](#)[About CE](#)[Contact CE](#)

Public Use Microdata (PUMD)

The Consumer Expenditure Survey (CE) collects data on expenditures, income, and demographics in the United States. The 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 a strong background with statistical software.

- [Public Use Microdata Files](#) - PUMD are available in three formats: SAS, STATA, and Comma Delimited (ASCII). Each format presents the data for each year in two zipped files. Data collected in the Interview survey are in the "Interview" file and those with the Diary Survey are in the "Diary" file.
- [CE PUMD Getting Started Guide](#) - This guide will help you understand the PUMD files and how to use them. It explains the methodology used in the CE survey and how to recreate the tables using PUMD.
- [Documentation](#) - Additional resources and files to help understand the PUMD.
- [Public Use Microdata Users' Workshop](#) - Provides users the opportunity to learn how to use the PUMD and to present how they used the PUMD in their research.

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](#).

Last Modified Date: May 4, 2023

RECOMMEND THIS PAGE USING: [Facebook](#) [Twitter](#) [LinkedIn](#)

<https://www.bls.gov/cex/pumd.htm>

Public Use Microdata

■ Data files:

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

■ PUMD Getting Start Guide:

- ▶ Help understanding the microdata structure, methodology, and how to use the data

■ Documentation:

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

■ Public Use Microdata Users' Workshop

PUMD Data Files

Public use microdata (PUMD) are available in three formats: SAS, 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.

Beginning in 2016, the program began providing a new research product, state weights for constructing state level estimates for California, Florida, New York, New Jersey, and Texas. New Jersey is available for 2016-2020, beginning in 2021 it was determined the sample for New Jersey was not large enough to support a state estimate. Users should take note that the state weights are considered a research product and may not be consistently available across the five listed states going forward as they are highly dependent on sample composition. [Documentation](#) with more information on state weights is also available, as well as [State Weights Tables](#).

Please note:

- Data for 2022 contain only the first two quarters (2 and 3 for Interview and 1 and 2 for Diary).
- Beginning with the 2020, PUMD packages for the Interview Survey only contain quarters 2, 3, and 4 of the year and quarter 1 of the following year. If you need quarter 1 of the the year listed, please download the previous year's data package.
- PUMD are not available for 1982 or 1983.
- SPSS format is no longer supported, SPSS users are encouraged to utilize the CSV files for analysis.

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 \(XLSX\)](#). If you plan to use CE income data in your analysis, the [User's Guide to Income Imputation in the CE](#) provides important information to know before starting. Additionally, we supply training programs and sample data ([ZIP](#)) and host a free [microdata data users' workshop](#).

If you are having difficulty downloading the data, please see the [instructions on how to download zip files](#).

	SAS	STATA	CSV
SAS			
2022	Interview (.zip) (23.7 MB) Diary (.zip) (5.31 MB)		
2021	Interview (.zip) (64.8 MB) Diary (.zip) (10.0 MB) State weights (.zip) (190 KB)		
2020	Interview (.zip) (63.4 MB) Diary (.zip) (9.0 MB) State weights (.zip) (163 KB)		
2019	Interview (.zip) (80.7 MB) Diary (.zip) (9.1 MB) State weights (.zip) (143 KB)		
2018	Interview (.zip) (86.9 MB) Diary (.zip) (9.6 MB) State weights (.zip) (150 KB)		
2017	Interview (.zip) (89.1 MB) Diary (.zip) (10.6 MB) State weights (.zip) (226 KB)		
2016	Interview (.zip) (91.6 MB) Diary (.zip) (10.5 MB) State weights (.zip) (114 KB)		
2015	Interview (.zip) (87.7 MB) Diary (.zip) (10.5 MB)		
2014	Interview (.zip) (99.1 MB) Diary (.zip) (12.3 MB)		
2013	Interview (.zip) (101.1 MB) Diary (.zip) (11.5 MB)		
2012	Interview (.zip) (106.7 MB) Diary (.zip) (14.2 MB)		

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



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?

State Weight Files

- State Weights are a research product
 - ▶ Do not meet OMB guideline, but still provide valuable information
- Weights added to the main data page
- Information on how to use the weights in the Getting Started Guide



Consumer Expenditure Surveys Public Use Microdata Getting Started Guide

This page provides documentation for Consumer Expenditure Surveys (CE) Public Use Microdata (PUMD), its conventions, files, sample code, and methodology.

Table of Contents

[Section 1. CE program](#)

[Section 2. CE PUMD](#)

[Section 3. Interview Survey](#)

[Section 4. Diary Survey](#)

[Section 5. Sample code](#)

[Section 6. Methodology](#)

[Section 7. Considerations](#)

Section 1. CE program

The Consumer Expenditure Surveys (CE) program provides data on expenditures, income, and demographic characteristics of consumers in the United States. The CE program provides these data in [tables](#), [LABSTAT databases](#), [news releases](#), [publications](#), and [public use microdata files](#).

CE data are collected by the Census Bureau for the Bureau of Labor Statistics (BLS) in two surveys, the Interview Survey for major and/or recurring items, and the Diary Survey for more minor or frequently purchased items. CE data are primarily used to revise the relative importance of goods and services in the market basket of the Consumer Price Index. The CE program conducts the only Federal household survey to provide information on the complete range of consumers' expenditures and income. For more information, see the [overview section](#) in the CE chapter in the BLS Handbook of Methods.

Section 2. CE PUMD

CE PUMD provide the individual responses to the two surveys from respondents. The data have been adjusted to protect the confidentiality of respondents. The CE PUMD allow researchers to analyze expenditure, income, and demographic data beyond what is provided in published tabulations.

2.1 CE PUMD files

CE PUMD include data from both the Interview and Diary Surveys. Most files are analogous between the two surveys; however, the Interview Survey files contain roughly 50 additional [detailed data files](#), as well as [paradata files](#) that provide detail about the collection process. All PUMD files cover a full calendar year. For years prior to 1996, file availability may be limited. [Table 3 Interview Survey files and content](#) lists the major files currently available, and their content. For a more comprehensive list of files provided in the CE PUMD, see the [Dictionary for the Interview and Diary Surveys](#).

www.bls.gov/cex/pumd-getting-started-guide.htm

Getting Started Guide

- Description and conventions of the data files
- Details information and files structure of the data files
- Provides methodology and sample code for the PUMD to recreate published tables
- Describes State Weight files and links to documentation
- Provides considerations that should be taken when using the PUMD

PUMD Documentation

This page contains the documentation for the CE public-use microdata (PUMD).

All year-specific Interview and Diary Survey documentation has been combined into the [PUMD Getting Started Guide](#).

- [PUMD Getting Started Guide](#) provides documentation for the CE PUMD, its conventions, files, and methodology.
- [Dictionary for Interview and Diary Surveys \(XLSX\)](#) provides variables and codes.
- [Sample code](#) provides code in SAS, STATA, and R to prepare annual averages using the data in the Interview Survey, Diary Survey, or both surveys (called "Integrated") and the results. The code may need to be adjusted for specific years.

Additional resources

- [Survey materials](#) lists the documents used to collect data for the Interview Survey and the Diary Survey.
 - [Hierarchical groupings \(zip\)](#) includes a description of each UCC along with its hierarchical standing within each expenditure or income category for a given year. Hierarchical groupings are not available for years prior to 1996. For years 1996 and forward, three file types are available:
 - **Integrated groupings** lists UCCs that the [CE tables](#) use, and identifies the survey source for the UCCs. These files use this naming convention: CE-HG-Integ-2017.
 - **Interview groupings** lists the UCCs from the Interview Survey. These files use this naming convention: CE-HG-Inter-2017. Not available for 1996.
 - **Diary groupings** list the UCCs from the Diary Survey. These files use this naming convention: CE-HG-Diary-2017. Not available for 1996.
- For each UCC, the files provide this information:

Column	Content	Start position	Options
1	Type of information in the line	1	1: First line of name 2: Second line of name *: Comment line
2	Level of aggregation	4	Highest level of aggregation is 1 and lowest 9.
3	Name of the UCC	7	
4	UCC lists the identifier of the UCC	70	
5	Source or purpose of the UCC	83	I: Interview survey D: Diary survey S: Statistical UCCs
6	Factor by which the mean has to be multiplied to match the annualized data in the published tables	86	1: Multiply times 1 4: Multiply times 4
			CUCHARS: CU characteristics FOOD: Food expenditures

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



Documentation

- PUMD Getting Started Guide
- Dictionary for Interview and Diary Survey (XLSX)
 - ▶ Provides variables and codes from 1996 forward.
- Hierarchical groupings
 - Establishes a relation between the published tables' series and their contributing expenditure classification codes (UCC).

Documentation Continued

- 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 background
- User's guide to income imputation (PDF)
 - ▶ Provides methodology

Other Documentation

- Protection of respondent confidentiality
 - ▶ Describes the recoding, topcoding, and suppression
- Errata
 - Provides information on errors and their correction to the files of a particular year.



Other Resources

- About CE
 - ▶ Events
 - ▶ CE Glossary
 - ▶ Information for Survey Participants
 - ▶ CE Questions and Answers



Questions and Answers

On this Page:

[Survey Overview](#)

[Definitions](#)

[Data Products](#)

[Data Adjustments](#)

[Survey Topics](#)

[Survey Design](#)

[Public Use Microdata \(PUMD\)](#)

[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?](#)
5. [How did the COVID-19 Pandemic impact the Consumer Expenditure Surveys?](#)

Data Products

6. [What types of data are available and in what form?](#)
7. [What is the most recent Annual Report about the Consumer Expenditure Surveys data?](#)
8. [Are historical data from the Consumer Expenditure Surveys available?](#)
9. [How are Consumer Expenditure data used to estimate experimental poverty thresholds?](#)
10. [How are the data from the Interview and Diary Surveys integrated?](#)
11. [Do the published data come from both surveys?](#)
12. [Are meals and rent as paid included in the CE program's definition of income?](#)

<https://www.bls.gov/cex/csxfaqs.htm#PUMD>



2022 PUMD Release

- In April we released the first half of 2022
 - ▶ Interview for Q2 and Q3
 - ▶ Diary for Q1 and Q2
- September 2022



Bryan Rigg

Section Chief, Public Use Microdata
Division of Consumer Expenditure Surveys

202-691-5495

Rigg.Bryan@bls.gov

