

Consumer Expenditure Surveys Public-Use Microdata (PUMD) Overview

Jimmy Choi

Division of Consumer Expenditure Surveys
CE Microdata User's Workshop
July 18-20, 2018



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.



Consumer Expenditure Survey

BROWSE CE

[CE HOME](#)[CE OVERVIEW](#)[CE NEWS RELEASES](#)[CE DATABASES](#)[CE TABLES](#)[CE PUBLIC-USE MICRODATA](#)[CE SURVEY FORMS](#)[CE PUBLICATIONS](#)[CE FAQs](#)[CE EVENTS IN YOUR AREA](#)[CONTACT CE](#)

SEARCH CE

[Go](#)

CE TOPICS

[REPORTS](#)[GEOGRAPHY](#)[GLOSSARY OF TERMS](#)[INFORMATION FOR CE RESPONDENTS](#)[POVERTY RESEARCH](#)

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. If you are new to CE PUMD data, you may want to explore the data availability with the **NEW** [Interactive CE Visualization Tool](#) and methodology with the [Getting Started Guide](#).

The [2015 public-use microdata](#) is the most recent and was released on August 30, 2016.

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](#).

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.

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

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 Survey

SHARE ON:     FONT SIZE:   PRINT: 

BROWSE CE

[CE HOME](#)[CE OVERVIEW](#)[CE NEWS RELEASES](#)[CE DATABASES](#)[CE TABLES](#)[CE PUBLIC-USE MICRODATA](#) ▶[CE SURVEY FORMS](#) ▶[CE PUBLICATIONS](#)[CE FAQs](#)[CE EVENTS IN YOUR AREA](#)[CONTACT CE](#)

SEARCH CE

CE TOPICS

[REPORTS](#) ▶[GEOGRAPHY](#) ▶[GLOSSARY OF TERMS](#)[INFORMATION FOR CE RESPONDENTS](#)[POVERTY RESEARCH](#)[DATA COMPARISONS](#)

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 CE PUMD data, you may want to explore the data availability with the **NEW** [Interactive CE Visualization Tool](#) and methodology with the [Getting Started Guide](#). 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](#)).

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

SAS

2015	Interview (zip) (87.7 MB)	Diary (zip) (10.5 MB)	Documentation
2014	Interview (zip) (99.1 MB)	Diary (zip) (12.3 MB)	Documentation
2013	Interview (zip) (101.1 MB)	Diary (zip) (11.5 MB)	Documentation
2012	Interview (zip) (106.7 MB)	Diary (zip) (14.2 MB)	Documentation

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?

Consumer Expenditure Surveys



BROWSE CE

- [CE HOME](#)
- [CE OVERVIEW](#)
- [CE NEWS RELEASES](#)
- [CE DATABASES](#)
- [CE TABLES](#)
- [CE PUBLIC USE MICRODATA](#)
- [CE SURVEY MATERIALS](#)
- [CE PUBLICATIONS](#)
- [CE FAQs](#)
- [CE EVENTS IN YOUR AREA](#)
- [CONTACT CE](#)

SEARCH CE



CE TOPICS

- [PUBLICATIONS](#) ▶
- [GEOGRAPHY](#) ▶
- [GLOSSARY OF TERMS](#)
- [INFORMATION FOR CE RESPONDENTS](#)
- [POVERTY RESEARCH](#)
- [DATA COMPARISONS](#)
- [CE REDESIGN](#)
- [METHODOLOGY](#)
- [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](#)
- » [Generational Research Tables](#)
- » [2014 Higher Income Table](#)
- » **NEW** [State Weights](#)
- » [All CU Prepublication MVP Tables](#)

How to get the most from CE tables

This [document](#) provides information to help one delve deeper into the published as well as the very detailed non-published (prepublication) tables. This short description describes how one can use consumer unit characteristics to uncover useful underlying information about the data.

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

Consumer Expenditure Survey

SHARE ON:

FONT SIZE:  PRINT: 

BROWSE CE

[CE HOME](#)[CE OVERVIEW](#)[CE NEWS RELEASES](#)[CE DATABASES](#)[CE TABLES](#)[CE PUBLIC-USE MICRODATA](#)[CE SURVEY FORMS](#)[CE PUBLICATIONS](#)[CE FAQs](#)[CE EVENTS IN YOUR AREA](#)[CONTACT CE](#)

SEARCH CE

Go

CE TOPICS

[REPORTS](#)[GEOGRAPHY](#)[GLOSSARY OF TERMS](#)[INFORMATION FOR CE RESPONDENTS](#)[POVERTY RESEARCH](#)[DATA COMPARISONS](#)[CE REDESIGN](#)[METHODOLOGY](#)[CE RESEARCH LIBRARY](#)[EXPERIMENTAL TABLES](#)

PUMD Documentation

This page contains documentation for the public-use microdata (PUMD) for years after 1996. Documentation for 1996 forward are below and for years prior on a USB flash drive for [purchase \(PDF\)](#). The documentation falls into two major types: [Documentation that covers all years](#) and [documentation that covers one particular year](#). If you are new to CE PUMD data, you may want to explore the data availability with the **NEW** [Interactive CE Visualization Tool](#) and methodology with the [Getting Started Guide](#).

Documents covering all years

- [Dictionary for Interview and Diary Survey \(XLSX\)](#) variables from 1996 forward.
- [Source selection file \(XLSX\)](#) identifies which survey data variable comes from when combining the two CE surveys for 1996 forward.
- [Response rate \(PDF\)](#) provides information on the response rates for the Interview and Diary survey for 2000 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

- **Documentation** of the content and methods for the data collected with the Interview Survey and Diary Survey. An asterisk after the word "Documentation", indicates that the documentation for the given year includes information on topcoding, response rates, errata, or other changes to the data.
- **Data dictionaries** list the variables, explanations, data flags, and data formats. The "Interview Survey dictionary" and "Diary Survey dictionary" are in PDF. The combined dictionary covers both surveys in an Access database. Before 2007, the dictionaries are included in the respective documentation.
- **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 Survey. Interview stub (IStub) lists the variables from the Interview Survey and Diary stub (DStub) from the Diary Survey. Before 2002, the documents contain a listing of Universal Classification Codes (UCCs), grouped by expenditure type, to define the summarized categories of the aggregated series in the published tables. The IntStub lists UCCs, along with an indicator identifying which survey (Interview or Diary) a value came from. The IStub lists UCCs from the Interview Survey and DStub from the Diary Survey. Before 2002, the documents contain an abbreviated hierarchy. PUSUMVARS provides a summary of key variables.

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

- Users' Documentation (Interview and Diary)
- Hierarchical groupings
- Sample code
- Errata



Users' Documentation

- Describes the content and methods for the data collected within the Interview and Diary Survey
- Includes definitions of our flags
 - Valid? Allocated? Imputed? Topcoded? Suppressed? A little of each?
- Estimation procedures/formulas
 - How to produce a collection year estimate
 - How to produce a calendar year estimate
- Information on changes to the data.
 - Variable changes
 - Questionnaire/Decennial Census sample redesign



Hierarchical Groupings

- Establishes a relation between the published tables' series and their contributing expenditure classification codes (UCC)
- Contains information on table aggregation, source selection, and UCC labels
- Three separate stub files are available
 - Interview Stub: IStubYYYY.txt
 - Diary Stub: DStubYYYY.txt
 - Integrated Stub: IntStubYYYY.txt



Hierarchical Groupings

1	2	Education	EDUCAT	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	Finance, late, interest charges for student loans	005520	I	4	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 2 school	660210	I	1	EXPEND
1	3	School books, supplies, equipment for vocational and 2 technical schools	660410	I	1	EXPEND
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
1	3	School supplies, etc. - unspecified	660000	D	1	EXPEND

IntStub2015, <https://www.bls.gov/cex/pumd/2015/csxintstub.txt>



Sample Programs

- SAS

- R

- STATA

- Goal is to replicate totals in the CE published tables
<https://www.bls.gov/cex/tables.htm>



Errata

- Provides information on errors and their correction to the files of a particular year.



[Home](#) ▾[Subjects](#) ▾[Data Tools](#) ▾[Publications](#) ▾[Economic Releases](#) ▾[Students](#) ▾[Beta](#) ▾

Consumer Expenditure Survey

SHARE ON:

FONT SIZE:  PRINT: 

BROWSE CE

[CE HOME](#)[CE OVERVIEW](#)[CE NEWS RELEASES](#)[CE DATABASES](#)[CE TABLES](#)[CE PUBLIC-USE MICRODATA](#)[CE SURVEY FORMS](#)[CE PUBLICATIONS](#)[CE FAQs](#)[CE EVENTS IN YOUR AREA](#)[CONTACT CE](#)

SEARCH CE



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.
- [How does the variability of Consumer Expenditure data impact your analysis?](#)

Last Modified Date: April 11, 2017

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

RECOMMEND THIS PAGE USING:



Facebook



Twitter



LinkedIn

Supplementary Resources

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



Consumer Expenditure Surveys

CE [SHARE ON:](#) [f](#) [t](#) [in](#) [PRINT:](#)

- BROWSE CE
- CE HOME
- CE OVERVIEW
- CE NEWS RELEASES
- CE DATABASES
- CE TABLES
- CE PUBLIC USE MICRODATA
- CE SURVEY MATERIALS
- CE PUBLICATIONS
- CE FAQs
- CE EVENTS IN YOUR AREA
- CONTACT CE

SEARCH CE [Go](#)

- CE TOPICS
- PUBLICATIONS
- GEOGRAPHY
- GLOSSARY OF TERMS

Frequently Asked Questions (FAQs)

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

- [What are the Consumer Expenditure Surveys?](#)
- [How are the Consumer Expenditure Surveys used?](#)
- [How do I contact the staff of the Consumer Expenditure Surveys?](#)
- [How do the Census Bureau and BLS handle respondent confidentiality?](#)

Data Products

- [What types of data are available and in what form?](#)
- [What is the most recent Annual Report about the Consumer Expenditure Surveys data?](#)

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

2017 PUMD Release

■ September 11th, 2018



Reminder!

- Microdata File Structure
 - ▶ Wednesday (7/18/18) at 9:45 a.m. - *Arcenis Rojas*

- Balancing Respondent Confidentiality and Data User Needs
 - ▶ Wednesday (7/18/18) at 1:00 p.m. - *Aaron Cobet*

- State Weights: Providing More Geographic Data
 - ▶ Thursday (7/19/18) at 9:40 a.m. - Taylor Wilson

- User's Guide to Income Imputation in the CE
 - ▶ Friday (7/20/18) at 1:30 p.m. - Geoffrey Paulin



Jimmy Choi

Economist

Division of Consumer Expenditure Surveys

202-691-7081

Choi.Jimmy@bls.gov

