The EEO Tabulation 2006-2010 (5-year ACS data)

FTP Site Technical Documentation

Table of Contents

Chapter 1 Introduction

Introduction

The Census Bureau entered into a reimbursable agreement with a consortium of four Federal agencies, consisting of the Equal Employment Opportunity Commission (EEOC), the Department of Justice (DOJ), the Office of Federal Contract Compliance Programs (OFCCP) at the Department of Labor (DOL), and the Office of Personnel Management (OPM), to create a custom tabulation identified as the EEO Tabulation 2006-2010 (5-year ACS data). This tabulation was created according to the specifications of the agencies in the consortium based on the American Community Survey (ACS) 2006-2010 5-year data. The EEO Tabulation 2006-2010 (5-year ACS data) serves as the primary external benchmark for comparing the race, ethnicity, and sex composition of an organization's internal workforce, and the analogous external labor market, within a specified geography and job category. More detailed information on this tabulation can be found here: http://www.census.gov/people/eeotabulation/.

The American Community Survey

The American Community Survey (ACS) is a part of the U.S. Census Bureau's Decennial Census Program and is designed to provide more current demographic, social, economic, and housing estimates throughout the decade. The ACS provides information on more than 40 topics, including education, language ability, the foreign-born, marital status, migration and many more subjects. Each year, the survey randomly samples around 3.5 million addresses and produces statistics that cover 1-year, 3-year, and 5-year periods for geographic areas in the United States and Puerto Rico. The 5-year estimates are available for many distinct geographies, including the nation, all 50 states, DC, Puerto Rico, counties, places, census tracts, and block groups. For more information about the ACS, please visit our home page at: www.census.gov/acs.

The 107 tables on the EEO Tabulation 2006-2010 (5-year ACS data) are available through American FactFinder (AFF) (factfinder2.census.gov). On AFF, these are available for download in several forms, including .csv files. This document will brief data users on the contents of the EEO tables located on the Census FTP site and explain how they can use it to obtain these tables.
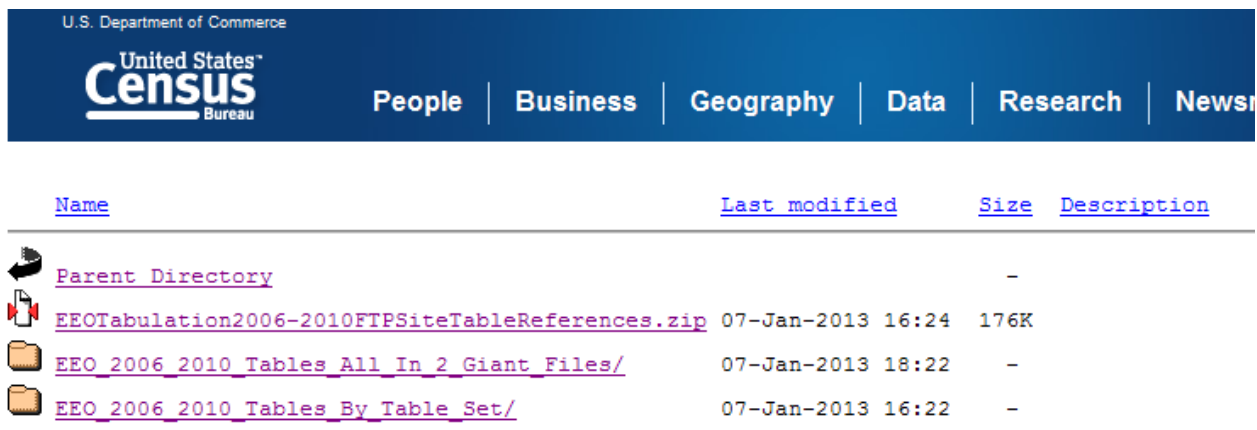
Chapter 2 How to Use the EEO Tabulation 2006-2010 (5-year ACS data) Files

2.1 Location of the EEO Tabulation (2006-2010) on the FTP site

The EEO Tabulation is located on the Census FTP site at
http://www2.census.gov/EEO_2006_2010/.

2.2 EEO Tabulation File Organization

The EEO Tabulation 2006-2010 (5-year ACS data) on the FTP site consists of a number of
zipped files. Each file set consists of one version of each table that covers all areas. The contents
of these files, how to use the datasets, and a description of the variables and geographies
contained in the datasets are included in this documentation.

U.S. Department of Commerce

United States Census Bureau

People | Business | Geography | Data | Research | Newsr

Name	Last modified	Size	Description
Parent Directory		–	
EEOTabulation2006-2010FTPSiteTableReferences.zip	07-Jan-2013 16:24	176K	
EEO_2006_2010_Tables_All_In_2_Giant_Files/	07-Jan-2013 18:22	–	
EEO_2006_2010_Tables_By_Table_Set/	07-Jan-2013 16:22	–	

The EEO Tabulation (2006-2010) on the FTP site is organized in two folders as shown in the
above screenshot. There are two directories that contain the same combination of files. These are
simply arranged differently to accommodate various user needs.

EEO_2006_2010_Tables_All_In_2_Giant_Files

The "Tables All in 2 Giant Files" directory contains two zipped files, which includes one
multiple data and annotation files divided by EEO Table sets 1 thru 7 and Table sets 8 thru 14.
These zipped files are ideal for users who want estimates and margins of error for all EEO
geographies throughout the nation at once. These files are very large and should only be used by
those that can easily process very large files.

EEO_2006_2010_Tables_By_Table _Set

The "Tables By Table Set" directory contains files for each table set for all EEO geographic
areas. Within those folders are geography files and files containing the data and annotation files,
one per table set (table sets are explained in Section 2.3). Downloading from these folders is

ideal for users who only want an entire table set for EEO geographies. These tables will be divided up by residence and workplace geographies.

The naming convention used for the zipped files in this directory is the following:

eeo-10_5yr-eeoall 1 r.zip

Program-
EEO Tabulation 2006-2010 (5-year ACS data)
Geography

r-Residence or w-Worksite

Table ID- 1 to 14

eeoall – All, Citizen, and not a citizen categories

Dataset name-
10_5yr- 2010 5-year estimates
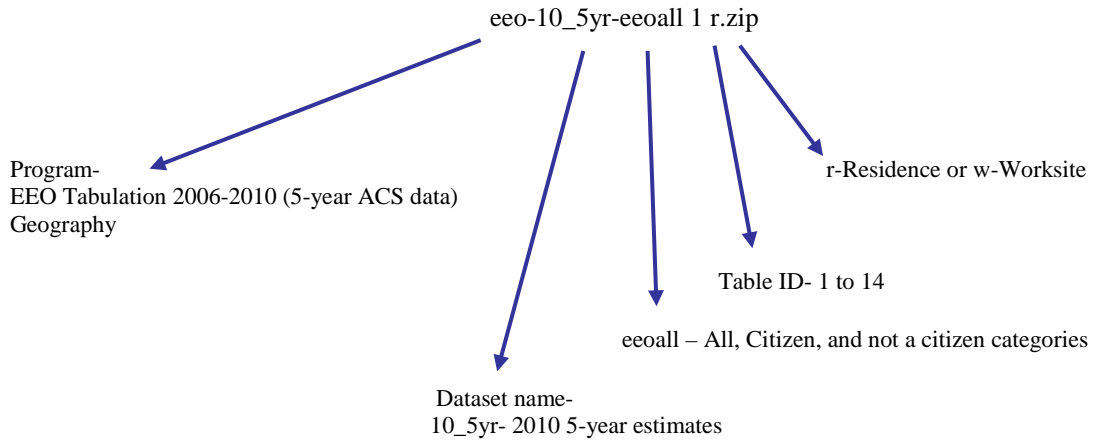
File Name: eeo-10_5yr-eeoall1r.zip		
Example	**Name**	**Range or Type**
eeo	Program	EEO Tabulation 2006-2010 (5-year ACS data)
10	Reference Year	ACS data year 2006-2010
5yr	Period Covered	5yr- 5-year data
eeoall	Categories Covered	Contains tables for categories of All (citizen and not a citizen combined), Citizen, and Not a citizen
1	Table ID	1 to 14
r/w	Geography Type	r-Residence or w-Worksite

2.3 Data Format and Access

ACSII text versions of the EEO Tabulation 2006-2010 (5-year ACS data) tables that are disseminated through AFF are available for downloading via FTP site. These files include the (1) geographic header record file, (2) data files and (3) annotation files in pipe-delimited ("|") ASCII format. These three types of files contain shared identifiers so these can be joined together. Metadata and EEO table shells are contained in a Microsoft Excel spreadsheet. More information about the files and metadata is described in detail in the sections below.

2.4 Data and Annotation Files

The data and annotation files are named <program>-< dataset>-<custom tab abbreviation>< table id><r-residence,w-worksite>.dat (or .ann). The program for this EEO Tabulation (2006-2010) is "eeo," the dataset will be the 2010 5-year data set (e.g.10_5yr), the tabulation abbreviation "eeo," the tableid is the EEO Table id (e.g.1 thru 14), followed by an "r" or "w"

denoting that the table set consists of residence or worksite geography. The file extension is .dat for the data files and .ann for the annotation files.

For example, .dat table and .ann table for EEO Table Set 1 for the EEO Tabulation 2006-2010 (5-year ACS data) for the total population in residence geographies are named as follows:

eeo-10_5yr-eeoall1r.dat
eeo-10_5yr-eeoall1r.ann

For the worksite geographies, these are named as follows:

eeo-10_5yr-eeoall1w.dat
eeo-10_5yr-eeoall1w.ann

2.5 Geographic Header File

Two geographic header files are provided which contain information about all geographic entities for the EEO Tabulation (2006-2010), one for the residence geographies and one for worksite/flow geographies. The geographic header file is named with the structure <program>-< dataset>-"-geographic-header_file-"-r/w.xls. The program for this tabulation is "eeo." The dataset will be the current 5-year data set (e.g. 10_5yr). The file extension is .xls. For example, the geographic header files for the EEO Tabulation 2006-2010 data are named as follows:

eeo-10-5yr-geographic-header-file-r.xlsx
eeo-10-5yr-geographic-header-file-w.xlsx

The header row in these files contains a record layout and each following row contains a single geographic area with a unique geographic identifier (GEOID). All geographic codes for a geographic area are embedded within the GEOID. The GEOID is made up of a three-digit summary level, a two-digit component (always "00" for EEO), a constant "US," and the unique geographic area code within the summary level. The summary level field (SUMLEV) is the critical element in identifying the geographic area type for each record. The worksite file contains a FLOWID and its associated GEOID. Both of these geographic header files are located in the zipped file eeo-10-5yr-geographic-header-file.zip on the EEO FTP site at http://www.census.gov/EEO_2006_2010/.

The SUMLEV field represents a three-digit code, which represent the following geographic areas:

- ➢ Nation (010)
- ➢ State (040)
- ➢ CBSA (310)
- ➢ State-County (050)
- ➢ State-Place (160)
- ➢ County Set (902)

2.6 Metadata File

The spreadsheet called "The EEO Tabulation 2006-2010-5-Year-Table-shells.xls" contains all of the table shells in a single location. See Table 2-1 below for an example of a table shell applied to the .dat structure and the section under File Structure called "Table Shell Metadata" for details about the metadata file.

Table 2-1. Example of an EEO table shell

AFF Display ID	Table ID	Indent	Line Num	Stub	Estimate_1	MOE_1
EEO-ALL01R	EEOALL1R	0	0	EEO 1r. Detailed Census Occupation by Sex and Race/Ethnicity for Residence Geography		
EEO-ALL01R	EEOALL1R	0	0.3			
EEO-ALL01R	EEOALL1R	0	0.4		Total, race and ethnicity	Total, race and ethnicity
EEO-ALL01R	EEOALL1R	1	0.9	Total, both sexes		
EEO-ALL01R	EEOALL1R	2	1	Number	#	#
EEO-ALL01R	EEOALL1R	2	2	Percent	%	%
EEO-ALL01R	EEOALL1R	1	2.9	Male		
EEO-ALL01R	EEOALL1R	2	3	Number	#	#
EEO-ALL01R	EEOALL1R	2	4	Percent	%	%
EEO-ALL01R	EEOALL1R	1	4.9	Female		
EEO-ALL01R	EEOALL1R	2	5	Number	#	#
EEO-ALL01R	EEOALL1R	2	6	Percent	%	%

2.7 Data File Contents

There are two types of tables that contain EEO Tabulation data. The data files (file extension - .dat) contain estimates and margins of error, but include numeric values for jam values (see Chapter 3 for more information on jam values). The annotation files (file extension - .ann) complement the .dat files by containing clearer character values for jam values.

NOTE: All of the following structural references to the data files also apply to the annotation files. In order to simplify documentation in this chapter, descriptions about the structure of the annotation files are omitted because they are redundant with the descriptions of the data files.

Each data file is a single EEO table in a pipe-delimited (|) ASCII file format. Each row contains the estimates and margins of error for a single geographic area for a particular iteration. The geographic area is identified with the first column, GEOID. Flow geographies contain a GEOID followed by a FLOW ID (i.e. 0500000US01001| 0100100000|) in the second field of the data file shown as GEOID|FLOWID|. The iteration number is the second field in the data file for non-flow data and indicates the EEO occupation iterations that are tabulated, and is in the third field for flow data. The iterations are identified by their four-digit iteration number (EEO OCC). See the example file layout for eeo-10_5yr-eeoall1w.dat in table 2-2 below for more details concerning the structure the data files.

Table 2-2. Example File Layout for eeo-10_5yr-eeoall1w.dat

```
GEOID             |FLOWID        OCC|EST1|MOE1|EST2|MOE2|……..
1600000US5613900|56139000000000|9750|21|32|0|119|0|119|35|42|0|222|0|222|0|222|9
1600000US7206593|72065937206593|9750|22|23|20|23|0|222|0|222|0|222|0|222|0|222|9
1600000US7214290|721429072119B0|9750|23|34|20|34|0|222|0|222|0|222|0|222|0|222|9
1600000US7232522|72325227276770|9750|9|24|4|24|0|222|0|123|0|222|0|222|0|222|0|0
1600000US7263820|72638207263820|9750|29|39|25|29|0|222|0|222|0|222|0|222|0|222|9
1600000US7276770|72767700000000|9750|210|202|210|202|0|219|0|219|0|219|0|219|0|9
1600000US7276770|72767707206593|9750|35|40|0|222|35|40|0|222|0|222|0|222|0|222|9
1600000US7276770|727677072137B0|9750|20|26|0|222|20|26|0|222|0|222|0|222|0|222|9
1600000US7276770|72767707276770|9750|70|53|45|46|20|23|0|222|15|25|0|222|0|222|9
```

The remaining fields in the data files alternate estimates and margins of error in numerical order of the table cells. The recommended naming for the estimates is tblid_tblcell_EST and for margins of error is tblid_ tblcell _ME. In most tables, field 3 contains the estimate for table cell 1 (i.e. eeoall1_1_EST, eeoall1_2_EST, etc.) and field 4 contains the margin of error for table cell 1 (i.e. eeoall1_1_ME, eeoall1_2_ME, etc.). If there is more than one cell in the table, then field 5 contains the estimate for table cell 2 and field 6 contains the margin of error for table cell 2, and so on for all table cells.

Chapter 3 User Notes

Supplemental Documentation

Supplemental documentation concerning the American Community Survey, to assist users using this technical document, is located on the ACS Website at: www.census.gov/acs/www/data_documentation/documentation_main/.

Documents such as the Subject Definitions, Accuracy of the Data, and Code Lists are available on the URL listed above.

Geographic Terms and Concepts

The most updated geographic terms and concepts can be found at: http://www.census.gov/geo/www/reference.html.

The EEO Tabulation 2006-2010 (5-year ACS data) uses the December 2009 vintage for metropolitan and micropolitan statistical areas and components which can be found at: http://www.census.gov/population/metro/files/lists/2009/List1.txt.

Data Collection and Processing Procedures

The American Community Survey operations involve a complex set of data collection and processing procedures that are too extensive to discuss as a chapter in the EEO Tabulation (2006-2010) documentation. EEO data users interested in a technical discussion of ACS operational processes and survey design should review the ACS Design and Methodology Report located at: http://www.census.gov/acs/www/methodology/methodology_main/.

Questionnaire

Examples of the questionnaires that were sent to sample addresses can be found in the questionnaires section of the ACS website located at: http://www.census.gov/acs/www/SBasics/SQuest/SQuest1.htm.

ACS Standard Data Products in New AFF and User Handbook

The U.S. Census Bureau produces a variety of products that allow data users to access and research ACS data. These products are available in American FactFinder (AFF). Information about how to use AFF is located at: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.
Data users looking for a summary overview of the ACS program or for discussion of annual changes in ACS data collection or product availability should refer to the Data and Documentation section on the ACS website at: http://www.census.gov/acs/www/data_documentation/data_main/.

Maps and Geographic Reference Materials

Detailed information about the ACS maps can be found in the reference maps section of the ACS website located at: http://www.census.gov/acs/www/data_documentation/reference_maps/.

Code Lists

The ACS provides code lists to identify all potential response categories for the variables included in the EEO Tabulation 2006-2010 (5-year ACS data). These variables are occupation, Hispanic origin, race, and industry. The code lists are available at: http://www.census.gov/acs/www/Downloads/data_documentation/pums/CodeLists/ACSPUMS2006_2010CodeLists.pdf.

Jam Values

Some data values represent unique situations where either the information to be conveyed is an explanation for the absence of data represented by a symbol in the data display, such as "(X). The data files (.dat) contain numeric values for jam values, while the annotation files (.ann) contain character values that are more descriptive.

The following list shows the special data values which can appear in any EEO Table on the FTP Site and on the American FactFinder website (AFF):

Special Data Values (Used in .dat files)	Display Value (Used in .ann files.)	Description
-999999999	N	Indicates that an estimate or its margin of error cannot be provided because the number of sample cases is too small for the given geographic area.
-888888888	(X)	Indicates that the estimate is not applicable or not available.
-666666666	-	Indicates that no sample observations were available to compute an estimate, or a ratio of medians cannot be calculated because one or both of the median estimates falls in the lowest interval or upper interval of an open-ended distribution.
-222222222	**	An '**' entry in the margin of error column indicates that either no sample observations or too few sample observations were available to compute a standard error and thus the margin of error. A statistical test is not appropriate.

A missing string indicates that the estimate is unavailable. This appears in the data files as two pipe-delimiters adjacent to each other without anything between them, or if the last cell in a data file is filtered then you get a pipe-delimiter followed immediately by a carriage return or EOF. A missing value indicates when an estimate is missing because of filtering for geographic restrictions or was removed due to the Disclosure Review Board's (DRB) Requirements.

Jam values are also used for the margins of error of controlled estimates. A statistical test for sampling variability is not appropriate. This is similar to the "*****" symbol used in American FactFinder.

Resources for using the FTP Site Files

The zipped file EEOTabulation2006-2010FTPSiteTableReferences.zip located on http://www2.census.gov/EEO_2006_2010/ contains the EEO-Tabulation-2006-2010-FTP-Site-Table-References.xlsx document, which consists of the following supplemental information for the EEO Tabulation 2006-2010 (5-year ACS data) disseminated via FTP.

- ➤ "**TableID-List**" defines the tables for EEO Tabulation (for FTP site).
- ➤ "**SummaryLevels-Table-Set**" defines the summary levels for EEO Special Tabulation.
- ➤ "**Tables-By-Geography**" defines all geographies used with which tables.
- ➤ "**Tables-By-Variables**" defines all variables used with which tables.
- ➤ "**Variables**" defines all variables that are referenced in the tables' lists, except for industry and for occupation.
- ➤ "**Industry**" defines the EEO 10W TableID suffix groupings, 2007 Census industry codes, corresponding 2007 NAICS industry codes, and EEO industry title.
- ➤ "**OccReCodesAggregatedOccs**" defines the EEO Tabulation 2006-2010 (5-year ACS) and 2008-2010 (3-year ACS) Occupation Code Crosswalk to Aggregated Occupations
- ➤ "**Guide-AggregatedOccups**" defines aggregated occupation group definitions.
- ➤ "**PUMSOCClist**" contains a crosswalk of full list of 2010 Census Detailed Occupation Codes to EEO Tabulation/PUMS Detailed Occupation Code List.
- ➤ "**EEOcrosswalk2000to0610**" contains an EEO Tabulation Crosswalk for the 2000 versus the 2006-2010 occupation codes.

User Notes

Other ACS user notes can be found on:
http://www.census.gov/acs/www/UseData/usernotes.html