# Urban Data Platform – Census/ACS Processing

# 2000, 2009, 2010, 2012, and 2015

*Last modified: CEO 5/18/2017*

## Overview

This document contains information on processing of all Census and American Community Survey data for the Urban Data Platform. All source data were downloaded from the Census Bureau from the links provided below. Data for all states, plus the District of Columbia and Puerto Rico, were downloaded and processed together. Then, data for the state of Texas, and select counties in Texas were extracted from the larger US file.

The years of Census or ACS data processed are as follows:

Census 2000 2000 Decennial Census of Population and Housing, 100% population counts

Census 2010 2010 Decennial Census of Population and Housing, 100% population counts

ACS 2009 American Community Survey 5-year estimate, covering years 2005-2009

ACS 2010 American Community Survey 5-year estimate, covering years 2006-2010

ACS 2012 American Community Survey 5-year estimate, covering years 2008-2012

ACS 2015 American Community Survey 5-year estimate, covering years 2011-2015

## Source Data

All raw source data were downloaded from the Census Bureau from the following:

|  |  |
| --- | --- |
| Census 2000 | <http://www2.census.gov/census_2000/datasets/Summary_File_1/>  <http://www2.census.gov/census_2000/datasets/Summary_File_3/> |
| ACS 2009 | <http://www2.census.gov/acs2009_5yr/summaryfile/> |
| Census 2010 | <http://www2.census.gov/census_2010/04-Summary_File_1/> |
| ACS 2010 | <http://www2.census.gov/acs2010_5yr/summaryfile/> |
| ACS 2012 | <http://www2.census.gov/acs2012_5yr/summaryfile/> |
| ACS 2015 | <http://www2.census.gov/programs-surveys/acs/summaryfile/2015/data/> |

These files contain raw counts for every available table, for every state, for all summary levels.

## Processing

The source files contain raw counts for every available table, for every state, for all summary levels. Additional documentation can be found in the technical documents supplied by the Census Bureau.

A series of SAS programs were run to extract tables of interest for all states for the following select summary levels: State, County, Tract, Block Group, Block, and ZCTA. Note that Block and ZCTA level data are not available from every source file. The SAS programs executed these standard processing steps:

1. Extract table data from source file – all states, all summary levels
2. Subset to include only select summary levels
3. Run internal checks on counts[[1]](#footnote-1)
4. Calculate standard counts
   1. For example, to calculate the standard age group 30-39, we sum the counts for males age 30-34, males age 35-39, females age 30-34, and females age 35-39.
5. Calculate standard percentages
6. Label standard variables
7. Output two files:
   1. A “raw” file contains all data elements from the source table, using the original source variable names. No calculations are done.
   2. A “standardized” file contains just the calculated standard variables and the geographic identifiers.
8. From the “standardized” file, the following geographic areas are extracted for the Urban Data Platform:
   1. The state of Texas
   2. 9 Counties in the Houston MSA: Austin, Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery and Waller
   3. Zip Code Tabulation Areas (ZCTAs)[[2]](#footnote-2)

See the companion Codebook file for the specific source table for each extracted/calculated standard variable.

## Published File(s), Programs and Audit Trail

Published datasets follow the following naming scheme:

[*source*][*year*][*estimat*e (ACS only)][*geo extent*]\_[*summary level*]\_v01

Where:

Source = Cen[[3]](#footnote-3), ACS

Year = 2000, 2009, 2010, 2012, 2015

(for 5-year estimates, 2009=2005-2009, 2010=2006-2010, 2012=2008-2012, and 2015=2011-2015)

Estimate = 5 (5-year ACS estimate)

Geo Extent = TX, Austin, Brazoria, Chambers, Fort\_Bend, Galveston, Harris, Liberty, Montgomery, or Waller

Summary Level = Blk, BG, Trt, Cnty, or ZCTA

v01 = Version 1 processing

All datasets are provided in the following formats: SAS (sas7bdat), Stata (.dta) .DBF, and .CSV.

Files for geographic extent=“TX”, and summary level=“Cnty” include an extra record that contains the state summary. For example, Cen2000TX\_Cnty\_v01 has 255 records – one for each county (254 counties) PLUS one record for the state summary.

Files for geographic extent=[County], and summary level=“Trt” include an extra record that contains the county summary. For example, Cen2000Austin\_Trt\_v01 has 6 records – one for each tract (5 tracts) PLUS one record for the Austin County summary.

|  |  |  |  |
| --- | --- | --- | --- |
| **Program Name** | **Input File(s)** | **Output File(s)** | **Date/Time Run** |
| Cen2000TX\_v01 | Cen2000US\_Blk\_std  Cen2000US\_BG\_std  Cen2000US\_Trt\_std  Cen2000US\_Cnty\_std | Cen2000TX\_Cnty\_v01  Cen2000TX\_\*\*\*\_v01  Cen2000Austin\_\*\*\*\_v01  Cen2000Brazoria\_\*\*\*\_v01  Cen2000Chambers\_\*\*\*\_v01  Cen2000Fort\_Bend\_\*\*\*\_v01  Cen2000Galveston\_\*\*\*\_v01  Cen2000Harris\_\*\*\*\_v01  Cen2000Liberty\_\*\*\*\_v01  Cen2000Montgomery\_\*\*\*\_v01  Cen2000Waller\_\*\*\*\_v01  \*\*\* = Blk, BG, or Trt | 10/20/2016  09:59:03 am |
| ACS20095TX\_v01 | ACS20095US\_BG\_ndises  ACS20095US\_Trt\_ndises  ACS20095US\_Cnty\_ndises | ACS20095TX\_Cnty\_v01  ACS20095TX\_\*\*\*\_v01  ACS20095Austin\_\*\*\*\_v01  ACS20095Brazoria\_\*\*\*\_v01  ACS20095Chambers\_\*\*\*\_v01  ACS20095Fort\_Bend\_\*\*\*\_v01  ACS20095Galveston\_\*\*\*\_v01  ACS20095Harris\_\*\*\*\_v01  ACS20095Liberty\_\*\*\*\_v01  ACS20095Montgomery\_\*\*\*\_v01  ACS20095Waller\_\*\*\*\_v01  \*\*\* is BG or Trt | 10/20/2016  02:10:32 pm |
| Cen2010TX\_v01 | Cen2010US\_Blk\_std  Cen2010US\_BG\_std  Cen2010US\_Trt\_std  Cen2010US\_Cnty\_std | Cen2010TX\_Cnty\_v01  Cen2010TX\_\*\*\*\_v01  Cen2010Austin\_\*\*\*\_v01  Cen2010Brazoria\_\*\*\*\_v01  Cen2010Chambers\_\*\*\*\_v01  Cen2010Fort\_Bend\_\*\*\*\_v01  Cen2010Galveston\_\*\*\*\_v01  Cen2010Harris\_\*\*\*\_v01  Cen2010Liberty\_\*\*\*\_v01  Cen2010Montgomery\_\*\*\*\_v01  Cen2010Waller\_\*\*\*\_v01  \*\*\* = Blk, BG, or Trt | 10/20/2016  10:02:49 am |
| ACS20105TX\_v01 | ACS20105US\_BG\_ndises  ACS20105US\_Trt\_ndises  ACS20105US\_Cnty\_ndises | ACS20105TX\_Cnty\_v01  ACS20105TX\_\*\*\*\_v01  ACS20105Austin\_\*\*\*\_v01  ACS20105Brazoria\_\*\*\*\_v01  ACS20105Chambers\_\*\*\*\_v01  ACS20105Fort\_Bend\_\*\*\*\_v01  ACS20105Galveston\_\*\*\*\_v01  ACS20105Harris\_\*\*\*\_v01  ACS20105Liberty\_\*\*\*\_v01  ACS20105Montgomery\_\*\*\*\_v01  ACS20105Waller\_\*\*\*\_v01  \*\*\* is BG or Trt | 10/20/2016  12:17:55 pm |
| ACS20125TX\_v01 | ACS20125US\_BG\_ndises  ACS20125US\_Trt\_ndises  ACS20125US\_Cnty\_ndises | ACS20125TX\_Cnty\_v01  ACS20125TX\_\*\*\*\_v01  ACS20125Austin\_\*\*\*\_v01  ACS20125Brazoria\_\*\*\*\_v01  ACS20125Chambers\_\*\*\*\_v01  ACS20125Fort\_Bend\_\*\*\*\_v01  ACS20125Galveston\_\*\*\*\_v01  ACS20125Harris\_\*\*\*\_v01  ACS20125Liberty\_\*\*\*\_v01  ACS20125Montgomery\_\*\*\*\_v01  ACS20125Waller\_\*\*\*\_v01  \*\*\* is BG or Trt | 10/20/2016  02:31:16 pm |
| ACS20155TX\_v01 | ACS20155US\_BG\_ndises  ACS20155US\_Trt\_ndises  ACS20155US\_Cnty\_ndises | ACS20155TX\_Cnty\_v01  ACS20155TX\_\*\*\*\_v01  ACS20155Austin\_\*\*\*\_v01  ACS20155Brazoria\_\*\*\*\_v01  ACS20155Chambers\_\*\*\*\_v01  ACS20155Fort\_Bend\_\*\*\*\_v01  ACS20155Galveston\_\*\*\*\_v01  ACS20155Harris\_\*\*\*\_v01  ACS20155Liberty\_\*\*\*\_v01  ACS20155Montgomery\_\*\*\*\_v01  ACS20155Waller\_\*\*\*\_v01  \*\*\* is BG or Trt | 05/11/2017  2:45:32 pm |

1. This is done to ensure that the correct table was extracted. Since the source files are just a long string of numbers, it would be easy to read in data “shifted” by one column. For example, if processing a table that includes Total Population, Total Male, and Total Female, an internal check can show that Total Male + Total Female = Total Population. If the data was read in incorrectly, this check will produce an error. [↑](#footnote-ref-1)
2. ZCTAs are available from CEHI, but have not yet been extracted for the UDP. [↑](#footnote-ref-2)
3. The Census 2010 files contain data from both the census and the ACS. These were combined by CEHI initially. In future versions, these should be separated. [↑](#footnote-ref-3)