Documentation for the
U.S. Census Grids, Summary File 1, 2010

April 2017

Center for International Earth Science Information Network (CIESIN),
Columbia University

Abstract

This document outlines the basic methodology and data sets used to construct the
Summary File 1, v1 (2010), along with use cases, limitations, and use constraints.

Data set citation:

Center for International Earth Science Information Network (CIESIN), Columbia
University. 2017. U.S. Census Grids (Summary File 1), 2010. Palisades, NY: NASA
Socioeconomic Data and Applications Center
(SEDAC). http://doi.org/10.7927/H40Z716C. Accessed DAY MONTH YEAR.

Suggested citation for this document:

Center for International Earth Science Information Network (CIESIN), Columbia
University. 2017. Documentation for the U.S. Census Grids (Summary File 1), 2010.
Palisades NY: NASA Socioeconomic Data and Applications Center
(SEDAC). http://doi.org/10.7927/H4MP517W. Accessed DAY MONTH YEAR.

We appreciate feedback regarding this data set, such as suggestions, discovery of errors,
difficulties in using the data, and format preferences. Please contact:

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
Columbia University
Phone: 1 (845) 365-8920
Email: ciesin.info@ciesin.columbia.edu

Contents

I.   Introduction

U.S. Census Grids, Summary File 1, 2010 contain grids of demographic and socioeconomic data from the 2010 U.S. Census. The data are available for download in ASCII, GeoTIFF, and Shapefile formats. The grids have a resolution of 30 arc-seconds (0.0083 decimal degrees), or approximately 1 square km at the equator. The gridded data are derived from census block geography from Census 2010 TIGER/Line Files, and variables on population, households, and housing from Census 2010 Summary File 1 (SF1). This data set is produced by the Columbia University Center for International Earth Science Information Network (CIESIN) using data provided by the U.S. Census Bureau.

II.   Data and Methodology

Input data

The gridded variables are based on census block geography from Census 2010 TIGER/Line Files, and variables on population, households, and housing from Census 2010 Summary File 1 (SF1).

TIGER/Line Files are "extracts of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) database". For a full technical description of TIGER/Line Files, see:

http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2010/TGRSHP10SF1.pdf

SF1 contains the data compiled from questions asked of all people and about every housing unit. U.S. Census Grids, Summary File 1, 2010 consists of information extracted from the sub-tables P1, P3, P4, P12, P14, P16, P18, P19, H1, H3, H14, H5, and H16 respectively. For a full technical description of SF1 see:

http://www.census.gov/prod/cen2010/doc/sf1.pdf

Methods

These 9 steps were used to develop this data set:

1. The 2010 SF1 files for each state were converted to text files and loaded into geodatabases.
2. The SF1 fields were combined as specified in the variable documentation to create the variables being gridded.
3. Non-Hispanic White (nhwhite) and Non-Hispanic Black (nhblack) were disaggregated from the tract level to the block level based on population distribution.
4. The 2010 TIGER/Line shapefiles were imported into a geodatabase.
5. The SF1 variables were joined to the Tiger feature classes. The density of each variable was calculated for each census block.
6. A 30 arc-second fishnet was created and intersected with the census block feature class.
7. The area in square kilometers was calculated for each area of intersection. The total count for each variable was calculated by multiplying the area in square kilometers in each area of intersection by the density of that variable in the underlying census block.
8. The total count for each grid cell was calculated by aggregating the variable counts for each area of intersection that falls within that grid cell.
9. The fishnet grid for each variable was converted to raster format.

III. Data Set Description(s)

Data set description:
Data are available at a resolution of 30 arc-seconds or approximately 1 kilometer for the continental United States, and separate files are available for Hawaii, Puerto Rico and Alaska. Summary File 1 (SF1) data are derived from the Census short form and are based on a full count of the population in 2010.

Data set web page:
http://www.sedac.ciesin.columbia.edu/data/set/usgrid-summary-file1-2010

Data set format:
The data are available in GeoTiff, ASCII, and Shapefile formats as downloadable zip files. Each downloadable is a compressed zip file, containing: 1) GeoTiff, ASCII, or Shapefile data, 2) land area grid, 3) PDF documentation, 4) data set variable name table [usgrid-2010-variables.xlsx], and 5) table containing state-level census statistics [usgrid-2010-states.xlsx].

Codebook:

Category	Description	File	Table	Field(s)	Grid abbreviation	Shapefile field name
Population	Population	SF1	P1	P0010001	pop	POP
Age	Population Under Age 1	SF1	P14	sum of P0140003 and P0140024	a1	AUND1
Age	Population Ages 1 to 4	SF1	P14	sum of P0140004 to P0140007 and P0140025 to P0140028	a2	A1TO4
Age	Population Ages 5 to 17	SF1	P14	sum of P0140008 to P0140020 and P0140029 to P0140041	a3	A5TO17
Age	Population Ages 18 to 24	SF1	P12	sum of P0120007 to P0120010 and P0120031 to P0120034	a4	A18TO24
Age	Population Under Age 25	SF1	P12, P14	sum of P0120007 to P0120010, P0120031 to P0120034, P0140003 to P0140020, and P0140024 to P0140041	u25	PUND25
Age	Population Age 25 and Older	SF1	P12	sum of P0120011 to P0120025 and P0120035 to P0120049	p25	P25
Age	Population Ages 25 to 64	SF1	P12	sum of P0120011 to P0120019 and P0120035 to P0120043	a5	A25TO64
Age	Population Ages 65 to 79	SF1	P12	sum of P0120020 to P0120023 and P0120044 to P0120047	a6	A65TO79
Age	Population Age 80 and Older	SF1	P12	sum of P0120024, P0120025, P0120048, and P0120049	a7	AOV80
Age	Women of Childbearing Age (Ages 15 to 49)	SF1	P12	sum of P0120030 to P0120039	wc	WOMCHILD
Race and Ethnicity	White Alone	SF1	P3	P0030002	wh	WHITE
Race and Ethnicity	Black or African American Alone	SF1	P3	P0030003	bl	BLACK
Race and Ethnicity	American Indian and Alaska Native alone	SF1	P3	P0030004	am	AMIND
Race and Ethnicity	Asian Alone	SF1	P3	P0030005	as	ASIAN

Race and Ethnicity	Native Hawaiian and Other Pacific Islander Alone	SF1	P3	P0030006	pi	HAWPI
Race and Ethnicity	Some Other Race Alone	SF1	P3	P0030007	or	OTHER
Race and Ethnicity	Two or More Races	SF1	P3	P0030008	two	TWOMORE
Race and Ethnicity	Not Hispanic	SF1	P4	P0040002	nhi	NHISP
Race and Ethnicity	Hispanic	SF1	P4	P0040003	hi	HISP
Race and Ethnicity	Non-Hispanic White	SF1	P4	PCT012I001 disaggregated from tract level to block	nhw	NHWHITE
Race and Ethnicity	Non-Hispanic Black	SF1	P4	PCT012J001 disaggregated from tract level to block	nhb	NHBLACK
Households	Households	SF1	P18	P0180001	hh	HH
Households	Population in Households	SF1	P16	P0160001	hhp	HHP
Households	Female Headed Households with Related Children Under Age 18	SF1	P19	P0190015	fem	FEM
Housing	Housing Units	SF1	H1	H00010001	hu	HU
Housing	Occupied Housing Units	SF1	H3	H0030001	occ	OCC
Housing	Owner Occupied Housing Units	SF1	H14	H0140002	own	OWN
Housing	Vacant Housing Unit, for Seasonal, Recreational or Occasional Use	SF1	H5	H0050006	sea	SEA
Housing	One-Person Housing Units	SF1	H16	sum of H0160003 and H0160011	hu1p	HU1P

IV.   How to Use the Data

The data can be used in any standard geographic information system (GIS) software package.

V.   Potential Use Cases

Gridded population data are useful for integration with other gridded data sets (e.g., from the earth and environmental sciences), and allow aggregation of population variables by geographies other than the original census units in which the census data are generally provided. Thus, for example, with this data set it is possible to easily sum the population

aged 0-15 by land use category, or to calculate the minority population within 10km of the ocean.

VI.   Limitations

The methodology applied for developing this data set is well established, and there are no known issues with this data set. Users are referred to the Census Bureau Web site and the peer reviewed literature for any potential issues with the census data (e.g., biases, undercounts, or the like).

VII.   Acknowledgments

This data set would have been far more difficult to produce without free and open access to data, especially the provision of census geographies, by the U.S. Census Bureau.

VIII. Disclaimer

CIESIN follows procedures designed to ensure that data disseminated by CIESIN are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact SEDAC User Services at ciesin.info@ciesin.columbia.edu. Neither CIESIN nor NASA verifies or guarantees the accuracy, reliability, or completeness of any data provided. CIESIN provides this data without warranty of any kind whatsoever, either expressed or implied. CIESIN shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided by CIESIN.

IX.   Use Constraints

X.   Recommended Citation(s)

Data set:

Center for International Earth Science Information Network (CIESIN), Columbia University. 2017. U.S. Census Grids (Summary File 1), 2010. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://doi.org/10.7927/H40Z716C. Accessed DAY MONTH YEAR.

XI.   Source Code

Please contact SEDAC User Services at ciesin.info@ciesin.columbia.edu to obtain the code written to produce this data set.

XII.   References

Input Data:
United States Census Bureau, 2010 Census, Summary File 1, http://www2.census.gov/census_2010/04-Summary_File_1, Accessed 05/01/2012.

United States Census Bureau, 2010 TIGER/Line Shapefiles, https://www.census.gov/geo/maps-data/data/tiger-line.html, Accessed 05/01/2012.

United States Census Bureau, Areal Hydrography, https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html, Accessed 05/01/2012.

Software:
Processed using Esri ArcGIS and Python.

The gridding method used was the same as in the Gridded Population of the World (GPW), v4 found here:
http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/methods/grid

XIII. Documentation Copyright and License

Copyright © 2017. The Trustees of Columbia University in the City of New York. This document is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

Appendix 1. Data Revision History

The U.S. Census Grids Summary File 1 (SF1) was first produced in 1990 and then again in 2000.

The U.S. Census Grids Summary File 1 (SF1) 2010 was produced in 2017 and includes:
- Several additional gridded variables that were not included in either 1990 or 2000 (Population Under Age 25, Women of Childbearing Age, Some Other Race Alone, Two or More Races, Not Hispanic)

Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	Contributors	Revisions
April 4, 2017	Jane Mills, Olena Borkovska, Tricia Chai Onn, Kytt MacManus	This document is the 1st instance of documentation