Documentation for the Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100

January 2021

Matt Hauer[1,2], Center for International Earth Science Information Network (CIESIN), Columbia University

[1,2] Department of Sociology, Florida State University, Tallahassee, USA. The Center for Demography and Population Health, Florida State University, Tallahassee, USA.

Abstract

This document outlines the basic methodology and data sets used for the development of the Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100, along with limitations, and use constraints. These data have numerous potential uses and can serve as inputs for addressing questions involving subnational demographic change in the United States.

Data set citation: Hauer, M., and Center for International Earth Science Information Network (CIESIN), Columbia University. 2021. Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/dv72-s254. Accessed DAY MONTH YEAR.

Suggested citation for this document: Center for International Earth Science Information Network (CIESIN), Columbia University. 2021. Documentation for the U.S. Georeferenced County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100. Palisades NY: NASA Socioeconomic Data and Applications Center. https://doi.org/10.7927/escf-qp41.

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

The Georeferenced U. S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100, consists of projections of total population, and population by sex, by age groups, and by race, in five-year intervals, for all U.S. counties for the period 2020 - 2100, in vector polygon and tabular formats.

The database builds on M. Hauer's (2019; 2019a) "Population projections for all U.S. counties by age, sex, and race controlled to the Shared Socioeconomic Pathways" (https://doi.org/10.17605/OSF.IO/9YNFC). This documentation includes the methods (including scripts[1]) used to transform and georeference the input data.

The purpose of the database is to provide small area and subnational population projections essential to understanding long-term demographic changes, planning for the near, medium and long term future, and making informed decisions in a variety of sectors. These data have numerous potential uses and provide a critical input with which to address questions of subnational demographic change in the United States under different future developmental and environmental scenarios.

II. Data and Methodology

This section details the sources and methods used to derive the counties' georeferenced layers and matching tables. Detailed information on the data sources and the

[1] STATA scripts are provided in Appendix 3.

methodology used for developing the tabular county population projections by age, sex and race, controlled to the SSPs, 2020-2100 (https://doi.org/10.17605/OSF.IO/9YNFC) is provided by Hauer (2019; 2019a).

Input data

The inputs for the total population projections table come from Hauer's (2019, 2019a) "Population projections for all U.S. counties by age, sex, and race controlled to the Shared Socioeconomic Pathways" (https://doi.org/10.17605/OSF.IO/9YNFC). The county digital geography is from the 2010 TIGER/Lines® Shapefiles (https://www2.census.gov/geo/tiger/TIGER2010/COUNTY/2010/).

Methods

1. Calculation of the total population projections for all counties, by SSP and 5-year intervals, 2020-2100

Calculations were made based on two steps. First, the input data were aggregated to create county population totals for each time interval and SSP scenario, by summing across categories of sex (2) race (4), and age (18), and obtaining a single total population value by county, SSP and year (Table 1.a). Second, the resulting database was transformed from long to wide format, so that each row represents a county, and the variables (columns) combine year and SSP scenario (Table 1.b)

Table 1. Data transformation from long to wide format

1.a. Long format of the input data aggregated to county's total population by year and SSP

year	geoid	ssp1	ssp2	ssp3	ssp4	ssp5
2020	**1001**	**55990**	**55835**	**54695**	**55490**	**57010**
2025	**1001**	**56230**	**55958**	**53641**	**55278**	**58310**

1.b. Wide format: county's total population values in a single row

geoid	ssp12020	ssp22020	ssp32020	ssp42020	ssp52020	ssp12025	ssp22025	ssp32025	ssp42025	ssp52025
01001	**55990**	**55835**	**54695**	**55490**	**57010**	**56230**	**55958**	**53641**	**55278**	**58310**

Scripts (in STATA) are included in Appendix 3.

2. Creation of the county total population projections vector polygon layer and table

The table with county population projection totals was joined to the 2010 TIGER Lines Shapefile using the unique identifier '*geoid*'. The resulting vector polygon layer displays total population projections by county, SSP scenario, and year in 5-year increments. The vector layer table is also available as a excel table.

3. Calculation of population projections aggregated by sex, race and age group, for all counties, by SSP and 5-year intervals, 2020-2100

These calculation follow the steps detailed in (1.) for each category of sex, race and age group. Scripts (STATA) are included in Appendix 3.

4. Creation of the vector polygon layers and tables for county population projections aggregated by sex, race and age group

The creation of these layers and tables follow the same steps detailed in (2.) above. There is a vector layer and table for each category of sex, race and age group.

A Data Dictionary defining all variables for total population, female and male population, Black NH (Non-Hispanic) population, Hispanic population, other population, White NH (Non-Hispanic) population, and across all age groups is available at:

- https://sedac.ciesin.columbia.edu/downloads/data/popdynamics/popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100/popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-data-dictionary-xlsx.zip

III. Data Set Description(s)

The Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100 provides population projections for all U.S. counties (total population, and population by sex, race and age) in 5-year intervals from 2020 to 2100. The data consists of:

1. 25 vector polygon layers:

- total population;
- female population; male population;
- Black NH (Non-Hispanic) population; Hispanic population; other population; White NH (Non-Hispanic) population;
- population age 0-4; population age 5-9; population age 10-14; population age 15-19; population age 20-24; population age 25-29; population age 30-34; population age 35-39; population age 40-44; population age 45-49; population age 50-54; population age 55-59; population age 60-64; population age 65-69; population age 70-74; population age 75-79; population age 80-84; population age 85+.

2. 25 tabular (excel) files matching the vector polygon layers.

Given that these data represent the United States, the North American Datum of 1983 (NAD 83) is the best projection for the population county data. In many other projections, areas like the Aleutian Islands appear disproportionate to their size or location. NAD 83 allows for visualization of the entire United States and corresponding territories in a more convenient manner. In addition, NAD 83 is compatible with data from SEDAC's U.S. Census Grids (Summary File 1), 2010 (CIESIN, 2017) and would therefore facilitate working with both data sets. For users interested in transforming the data, however, a simple Geographic Transformation can be conducted using ArcMap or QGIS, as described below.

In order to transform the coordinate system using ArcGIS: add the shapefiles to your ArcMap document; in the Geoprocessing toolbox, select the "Project" Data Management tool; enter the shapefile that you want to reproject and then select the new coordinate system as well as the transformation type; and then save the new shapefile, which will now be in the new coordinate system. When measuring distances and areas, you will likely want a projected coordinate system, which allows for accurate representation of the distance between places. Geographic coordinate systems are useful for visualizing places on Earth, but measurement units are not uniform and measurements are not reliable. For more information on choosing an appropriate transformation in ArcGIS, see Esri's help pages: https://desktop.arcgis.com/en/arcmap/latest/map/projections/choosing-an-appropriate-transformation.htm.

In order to transform the coordinate system using QGIS, you will need to export a new file, although enabling "on the fly" projection using the Coordinate Reference System (CRS) Status button in the Status Bar at the bottom of the QGIS window will allow you to temporarily reproject the layers on the map. You can save your layer in the coordinate system of your choice by right clicking the layer, selecting "Save as," followed by the "Browse" button next to "Save vector layer as." Specify the name of the new layer and change the value of the "Layer CRS" to "Selected CRS." The "Browse" button below the dropdown will reveal the CRS Selector. Choose the coordinate system of your choice, leaving "Symbology export" unchanged, and now save the vector layer.

Data set web page:
SEDAC URL: https://sedac.ciesin.columbia.edu/data/set/popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100

Permanent URL: https://doi.org/10.7927/dv72-s254

Data set format:
The data are available as vector polygon files in Shapefile (SHP) and as tables in Microsoft Excel (XLSX) format. The downloadables are compressed zip files containing: 1) SHP or XLSX, 2) Data Dictionary, and 3) PDF documentation.

Data set downloads:
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-pop-age-groups-0-25-shp.zip
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-pop-age-groups-30-55-shp.zip
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-pop-age-groups-60-85over-shp.zip
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-pop-total-sex-race-shp.zip
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-pop-total-sex-race-age-groups-xlsx.zip
popdynamics-us-county-level-pop-projections-sex-race-age-ssp-2020-2100-data-dictionary-xlsx.zip

IV. How to Use the Data

The vector data in Shapefile and tabular data in Microsoft Excel formats can be used directly in mapping and geospatial analysis.

V. Potential Use Cases

Potential uses of the data include calculating future scenarios of disease incidence (Gorris et al., 2019; Rohat 2020), current and future environmental vulnerability assessments (Keenan and Hauer 2020; Rohat 2019), and changes in consumption patterns (Obringer et al., 2020).

Predictions of future disease spread rely on population density and dynamics in order to calculate current and future spread. In documenting the future spread of Valley Fever, Gorris et al., tracked how many people live in regions exposed to the fungus as well as climate patterns relevant to its dispersal. In their 2020 analysis of socio-economic vulnerability to the effects of climate change on disease spread, Rohat et al., population data were critical to producing predicted future county-level exposure rates.

Environmental vulnerability analyses similarly rely upon accurate demographic information to assess the scale of impact on people in a given area. A 2019 study by Rohat et al., draws on Hauer’s (2019; 2019a) population projections data to evaluate extreme heat vulnerability in urban areas.

Georeferenced population data are critical towards understanding, for example, hotspots for heat exposure and the precise areas where adaptation and/or additional resources are needed. These data also enable greater analyses of demographic dynamics, vulnerabilities, and future patterns, as well as the potential of being applied to a broad set of environmental, economic, epidemiological, or other studies.

VI. Limitations

Limitations of the input data set are addressed in Hauer (2019:13) as follows:

"These projections, like all projections, involve the use of assumptions about future events that may or may not occur. Users of these projections should be aware that although the projections have been prepared with the use of standard methodologies, documentation of their creation, open-source computer code, and extensive evaluations of their accuracy and uncertainty, they might not accurately project the future population of a state, county, age, sex, or race group. The projections are based on historical trends and current estimates. Any small error in the projections early in the projection horizon could cascade into considerable errors decades later in the projection. Caveat emptor – users beware. These projections should be used only with full awareness of the inherent limitations of population projections in general and with knowledge of the procedures and assumptions described in this document"

In addition, Striessing et al., (2019) and Jiang et al., (2020) have suggested that the methodology used in the development of the input data (Hauer 2019; 2019a) could be more appropriate for shorter time periods.

VII. Acknowledgments

M. Hauer contributed the original input data (documented in Hauer 2019; 2019a). CIESIN developed the georeferenced data set described in this documentation.

Funding for development and dissemination of this data set was provided under the U.S. National Aeronautics and Space Administration (NASA) contract 80GSFC18C0111 for the continued operation of the Socioeconomic Data and Applications Center (SEDAC), which is operated by the Center for International Earth Science Information Network (CIESIN) of Columbia University.

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

This work is licensed under the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0).

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

X. Recommended Citation(s)

Data set(s):

Hauer, M., and Center for International Earth Science Information Network (CIESIN), Columbia University. 2021. Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/dv72-s254. Accessed DAY MONTH YEAR.

Scientific publication:

Hauer, M. E. 2019. Population Projections for U.S. Counties by Age, Sex, and Race Controlled to Shared Socioeconomic Pathway. Scientific Data 6: 190005. https://doi.org/10.1038/sdata.2019.5.

XI. Source Code

STATA scripts used to transform and georeference the input data are provided in Appendix 3.

XII. References

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). https://doi.org/10.7927/H40Z716C.

Gorris, M. E., et al. 2019. Expansion of Coccidioidomycosis Endemic Regions in the United States in Response to Climate Change. *Geohealth* **3**(10):308-327. https://doi.org/10.1029/2019GH000209.

Hauer, M. E. 2019. Population Projections for U.S. Counties by Age, Sex, and Race Controlled to Shared Socioeconomic Pathway. *Scientific Data*, **6**:190005. https://doi.org/10.1038/sdata.2019.5.

Hauer, M. 2019a. Population projections for all U.S. counties by age, sex, and race controlled to the Shared Socioeconomic Pathways. https://doi.org/10.17605/OSF.IO/9YNFC.

Jiang, L., B. C. O'Neill, H. Zoraghein, and S. Dahlke. 2020. Population scenarios for US states consistent with shared socioeconomic pathways. *Environmental Research Letters*, 15(9), 094097. https://doi.org/10.1088/1748-9326/aba5b1.

Keenan, J. M. and M. E. Hauer. 2020. Resilience for whom? Demographic change and the redevelopment of the built environment in Puerto Rico. *Environmental Research Letters*, **15**(7). https://doi.org/10.1088/1748-9326/ab92c2.

Obringer, R., et al. 2020. Managing the water-electricity demand nexus in a warming climate. *Climatic Change*, **159**(2): 233-252. https://doi.org/10.1007/s10584-020-02669-7.

Rohat, G., et al. 2020. Intersecting vulnerabilities: climatic and demographic contributions to future population exposure to Aedes-borne viruses in the United States. *Environmental Research Letters*, **15**(8). https://doi.org/10.1088/1748-9326/ab9141.

Striessnig, E., J. Gao, B. C. O'Neill, and L. Jiang. 2019. Empirically based spatial projections of US population age structure consistent with the shared socioeconomic pathways. *Environmental Research Letters*, 14(11), 114038. https://doi.org/10.1088/1748-9326/ab4a3a.

U.S. Census Bureau. 2012. 2010 TIGER/Line Shapefiles [machine-readable data files]. Washington, D.C. U.S. Census Bureau. https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010&layergroup=Counties+%28and+equivalent%29

U.S. Census Bureau. 2012b. TIGER/Line Shapefiles 2010. Technical Documentation. Washington, D.C. U.S. Census Bureau. https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2010/TGRSHP10SF1.pdf.

XIII. Documentation Copyright and License

Appendix 1. Data Revision History

No revisions have been made to this data set.

Appendix 2. Contributing Authors & Documentation Revision History

Revision Date	Contributors	Revisions
January 28, 2021	Susana B. Adamo, Anne-Laure White	This document is the 1st instance of documentation

Appendix 3. Data Transformation Scripts (STATA)

Total population projections

```
use
*open Hauer's dataset from https://doi.org/10.17605/OSF.IO/9YNFC*
collapse (first) state county (sum) ssp1 ssp2 ssp3 ssp4 ssp5, by (year geoid)
save
*save collapsed total population data in your files*

Use
*open collapsed total population data file*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped total population data in your files*
```

Population projections by sex

```
use
*open Hauer's dataset from https://doi.org/10.17605/OSF.IO/9YNFC*
collapse (first) state county (sum) ssp1 ssp2 ssp3 ssp4 ssp5, by (sex year
geoid)
save
*save collapsed population by sex data in your files*
```

Male population

```
Use
*open collapsed population by sex file*

keep if sex==1
save
*save collapsed population by sex/male data in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by sex/male data in your files*
```

Female population

```
Use
*open collapsed population by sex file*

keep if sex==2
save
```

```
*save collapsed population by sex/female data in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by sex/female data in your files*
```

Population projections by race

```
Use
*open Hauer's dataset from https://doi.org/10.17605/OSF.IO/9YNFC*
collapse (first) state county (sum) ssp1 ssp2 ssp3 ssp4 ssp5, by (race year
geoid)
save
*save collapsed population by race in your files*
```

White NH

```
Use
*open collapsed population by race file*
keep if race==1
save
*save collapsed population by race/White in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by race/White in your files*
```

Black NH

```
use
*open collapsed population by race file*
keep if race==2
save
*save collapsed population by race/Black in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by race/Black in your files*
```

Hispanic

```
use
*open collapsed population by race file*
keep if race==3
save
*save collapsed population by race/Hispanic in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by race/Hispanic in your files*
```

Other

```
use
*open collapsed population by race file*
keep if race==4
save
*save collapsed population by race/Other in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by race/Other in your files*
```

Population projections by age-group

```
use
*open Hauer's dataset from https://doi.org/10.17605/OSF.IO/9YNFC*
collapse (first) state county (sum) ssp1 ssp2 ssp3 ssp4 ssp5, by (age year
geoid)
save
```

```
*save collapsed population by age in your files*
```

Age groups

```
use
*open collapsed population by age file*
keep if age==1
save
*save collapsed population by age/1 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/1 in your files*

use
*open collapsed population by age file*
keep if age==2
save
*save collapsed population by age/2 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/2 in your files*

use
*open collapsed population by age file*
keep if age==3
save
*save collapsed population by age/3 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/3 in your files*

use
*open collapsed population by age file*
keep if age==4
save
*save collapsed population by age/4 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/4 in your files*

use use
*open collapsed population by age file*
keep if age==5
save
*save collapsed population by age/5 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/5 in your files*

use
*open collapsed population by age file*
keep if age==6
save
*save collapsed population by age/6 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/6 in your files*

use
*open collapsed population by age file*
keep if age==7
save
*save collapsed population by age/7 in your files*
```

```
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/7 in your files*

use
*open collapsed population by age file*
keep if age==8
save
*save collapsed population by age/8 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/8 in your files*

use
*open collapsed population by age file*
keep if age==9
save
*save collapsed population by age/9 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/9 in your files*

use
*open collapsed population by age file*
keep if age==10
save
*save collapsed population by age/10 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/10 in your files*

use
*open collapsed population by age file*
keep if age==11
save
*save collapsed population by age/11 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/11 in your files*

use
*open collapsed population by age file*
keep if age==12
save
*save collapsed population by age/12 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/12 in your files*

use
*open collapsed population by age file*
keep if age==13
save
*save collapsed population by age/13 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/13 in your files*

use
*open collapsed population by age file*
keep if age==14
save
```

```
*save collapsed population by age/14 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/14 in your files*

use
*open collapsed population by age file*
keep if age==15
save
*save collapsed population by age/15 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/15 in your files*

use
*open collapsed population by age file*
keep if age==16
save
*save collapsed population by age/16 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/16 in your files*
use
*open collapsed population by age file*
keep if age==17
save
*save collapsed population by age/17 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/17 in your files*

use
*open collapsed population by age file*
keep if age==18
save
*save collapsed population by age/18 in your files*
reshape wide ssp1 ssp2 ssp3 ssp4 ssp5, i(geoid) j(year)
save
*save reshaped population by age/18 in your files*
```