Documentation for the
U.S. Social Vulnerability Index Grids

August 2021

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

Abstract

The U.S. Social Vulnerability Index Grids data set contains gridded layers for the overall Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) using four sub-category themes (Socioeconomic, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation) based on census tract level inputs from 15 variables for the years 2000, 2010, 2014, 2016, and 2018. SVI values range between 0 and 1 based on their percentile position among all census tracts in the U.S., with 0 representing lowest vulnerability census tracts and 1 representing highest vulnerability census tracts. SEDAC has gridded these vector inputs to create 1 km spatial resolution raster surfaces allowing users to obtain vulnerability metrics for any user-defined area within the U.S. Utilizing inputs from CIESIN's Gridded Population of the World, Version 4 (GPWv4) Revision 11 data sets, a mask is applied for water, and optionally, for no population. The data are provided in two different projection formats, NAD83 as a U.S. specific standard, and WGS84 as a global standard. The goal of the SVI is to help identify vulnerable communities by ranking them on these inputs across the U.S.

Data set citation: Center for International Earth Science Information Network, (CIESIN), Columbia University. 2021. U.S. Social Vulnerability Index Grids. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/6s2a-9r49. 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. Social Vulnerability Index Grids. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/fjr9-a973. 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

Documentation last revised August 24, 2021.

Contents

I.   Introduction

Vulnerability is the "propensity or predisposition to be adversely affected" (IPCC, 2012). Vulnerability cannot be measured directly, therefore, proxies are constructed through indices. Typically, vulnerability was considered in relation to specific hazards, however, the Social Vulnerability Index (SVI) was developed to serve as a proxy of the vulnerability of the population in the United States. The Centers for Disease Control and Prevention (CDC) developed this Social Vulnerability Index (SVI) using an additive equal weighting approach based on aggregate census data for both the county and census tract levels (CDC, 2021). Social vulnerability indices are intended to indicate areas of relatively higher social disparities which can highlight areas with greater vulnerability (Cutter, 2005). The aim is to identify vulnerable communities which may require additional support in preparing for hazards or recovering from disaster (CDC, 2015).

CDC uses 15 census variables at the census tract level to produce an SVI score, which ranges between 0 and 1, to indicate relatively low or high vulnerability, respectively. The SEDAC Social Vulnerability Index Grids data set rasterizes these census tract-level social vulnerability measures at 1 km spatial resolution to fit the standard grid of CIESIN's Gridded Population of the World, Version 4 (GPWv4) Revision 11 and U.S. Census Grid families of population products. In addition to the general SVI indicator,

gridded layers are provided for four sub-category themes (Socioeconomic, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation). SEDAC gridded the vector SVI data for each of the five layers for each of the years 2000, 2010, 2014, 2016, and 2018.

II.   Data and Methodology

Input data

CDC's Social Vulnerability Index uses 15 variables at the census tract level. The data comes from the U.S. decennial census for the years 2000 & 2010, and the American Community Survey (ACS) for the years 2014, 2016, and 2018. It is a hierarchical additive index (Tate, 2013), with the component elements of CDC's SVI including the following for 4 themes: Socioeconomic Status (Below Poverty, Unemployed, Income, No High School Diploma), Household Composition & Disability (Aged 65 or Older, Aged 17 or Younger, Civilian with a Disability, Single-Parent Households), Minority Status & Language (Minority, Speaks English "Less than Well"), and Housing Type & Transportation (Multi-Unit Structures, Mobile Homes, Crowding, No Vehicle, Group Quarters). While state ranked versions of the indices are available, note that all of these data which are gridded are from the national U.S. data set version.

In regards to spatial formats, SVI 2000 and 2010 are downloadable in Esri Geodatabase format while 2014, 2016, and 2018 are downloadable from the CDC website as shapefiles. Each year the CDC SVI data contains a different number of census tracts: 2000 (65,081), 2010 (72,891), 2014 (72,842), 2016 (72,836), and 2018 (72,837). The SVI polygon data from the CDC are produced using the EPSG:4269 NAD83 projection and the processing is done in this projection. The original NAD83 layer is provided along with a layer reprojected to EPSG:4326 WGS84 to match the standard GPW grid.

CIESIN's Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11 (CIESIN, 2018a) data set provided the water mask while the Population Count, Revision 11 (CIESIN, 2018b) data set provided an optional population mask. The resolution of this data set is 30 arc-seconds or approximately 1 km at the equator.

Methods

SEDAC grids the CDC's pre-existing SVI layer in order to produce overall SVI scores as well as scores for each sub-category theme at 1 km resolution for the contiguous U.S. SVI data at the census tract level were downloaded from the CDC website for the years 2000, 2010, 2014, 2016, and 2018. The first two files of 2000 and 2010 are converted from Esri Geodatabase files (.mdb) to shapefiles while later years are directly downloadable as shapefiles. SEDAC uses an R script to grid the vector SVI value data and fit the polygon values to the standard grid of the GPWv4.11 (30 arc-seconds) population products. The SVI grid data were generated using a rasterization of polygons

approach using the center of each grid cell as input, and if this central point is not available, it is drawn from the value of overlapping polygons. The SVI values are evenly allocated within the tract as there were no finer grained measures to use at this time to allocate or weight the spatial distribution of social vulnerability within tracts. This creates a 1 km gridded raster surface that visually matches the polygon census tract units.

The raster layers store the SVI values which range of 0 to 1. In the SVI, low values represent low vulnerability while high values represent high vulnerability. The GPWv4: Land and Water Area, Revision 11 (CIESIN, 2018a) data were used to mask out grid cells that are water. This creates a no data value in our product for pixels that do not contain land. No data values also result from grid cells that are without input SVI data which varies slightly by the census year. Finally, a layer option is provided that is the same as described above, but adds an additional mask using GPWv4: Population Count, Revision 11 (CIESIN, 2018b) data of areas of no population.

This process was completed for the years 2000, 2010, 2014, 2016, and 2018 and for the five layers including the overall SVI value and the four sub-category themes of Socioeconomic, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation. Finally, a quality assurance check was performed for each layer with random checks of pixels from the rasters to confirm the polygon data was appropriately converted for each raster layer.

The SVI vector data columns that are formed into raster layers are included below in Table 1 to clearly indicate the source data of each output GeoTIFF layer:

Table 1

Originator Data Set/Year	Spatial Format	Original column name Documentation description	Output raster layer
CDC SVI 2000	.mdb	**USTP** Total Percentile Ranking	svi_2000_tract_overall.tif
CDC SVI 2000	.mdb	**USG1TP** Socioeconomic Domain Total Percentile Ranking	svi_2000_tract_socioeconomic.tif
CDC SVI 2000	.mdb	**USG2TP** Household Composition & Disability Total Percentile Ranking	svi_2000_tract_household.tif
CDC SVI 2000	.mdb	**USG3TP** Minority Status/Language Domain Total Percentile Ranking	svi_2000_tract_minority.tif
CDC SVI 2000	.mdb	**USG4TP** Housing/Transportation Domain Total Percentile Ranking	svi_2000_tract_housing.tif
CDC SVI 2010	.mdb	**R_PL_THEME** Overall Percentile Ranking (documentation: R_PL_THEMES)	svi_2010_tract_overall.tif
CDC SVI 2010	.mdb	**R_PL_THEME1** Percentile ranking for Socioeconomic theme	svi_2010_tract_socioeconomic.tif
CDC SVI 2010	.mdb	**R_PL_THEME2** Percentile ranking for Household Composition theme	svi_2010_tract_household.tif
CDC SVI 2010	.mdb	**R_PL_THEME3** Percentile ranking for Minority Status/Language theme	svi_2010_tract_minority.tif
CDC SVI 2010	.mdb	**R_PL_THEME4** Percentile ranking for Housing Type/Transportation theme	svi_2010_tract_housing.tif
CDC SVI 2014	.shp	**RPL_THEMES** Overall percentile ranking	svi_2014_tract_overall.tif
CDC SVI 2014	.shp	**RPL_THEME1** Percentile ranking for Socioeconomic theme	svi_2014_tract_socioeconomic.tif

CDC SVI 2014	.shp	**RPL_THEME2** Percentile ranking for Household Composition theme	svi_2014_tract_household.tif
CDC SVI 2014	.shp	**RPL_THEME3** Percentile ranking for Minority Status/Language theme	svi_2014_tract_minority.tif
CDC SVI 2014	.shp	**RPL_THEME4** Percentile ranking for Housing Type/Transportation theme	svi_2014_tract_housing.tif
CDC SVI 2016	.shp	**RPL_THEMES** Overall percentile ranking	svi_2016_tract_overall.tif
CDC SVI 2016	.shp	**RPL_THEME1** Percentile ranking for Socioeconomic theme summary	svi_2016_tract_socioeconomic.tif
CDC SVI 2016	.shp	**RPL_THEME2** Percentile ranking for Household Composition theme	svi_2016_tract_household.tif
CDC SVI 2016	.shp	**RPL_THEME3** Percentile ranking for Minority Status/Language them	svi_2016_tract_minority.tif
CDC SVI 2016	.shp	**RPL_THEME4** Percentile ranking for Housing Type/Transportation them	svi_2016_tract_housing.tif
CDC SVI 2018	.shp	**RPL_THEMES** Overall percentile ranking	svi_2018_tract_overall.tif
CDC SVI 2018	.shp	**RPL_THEME1** Percentile ranking for Socioeconomic theme summary	svi_2018_tract_socioeconomic.tif
CDC SVI 2018	.shp	**RPL_THEME2** Percentile ranking for Household Composition theme	svi_2018_tract_household.tif
CDC SVI 2018	.shp	**RPL_THEME3** Percentile ranking for Minority Status/Language them	svi_2018_tract_minority.tif
CDC SVI 2018	.shp	**RPL_THEME4** Percentile ranking for Housing Type/Transportation them	svi_2018_tract_housing.tif

For more information, see Appendix 3: Data Files

The resulting rasters contained gridded Social Vulnerability Index values at a 1 km fit to the GPW grid. The method was repeated to produce 5 layers for each of the 5 years of data with each year having one layer representing overall SVI and 4 sub-themes including Socioeconomic, Household Composition, Minority Status/Language, and Housing/Transportation. An additional projection specification is added to the end of each layer name to indicate either NAD83 or WGS84. Finally, "NoPop" is indicated for rasters with masks for water and no population applied. The 1 km raster grid data are downloadable in GeoTIFF format categorized by year and projection.

III.   Data Set Description(s)

The Social Vulnerability Index Grids data set represents social vulnerability at 1 km spatial resolution. Grid cell values range from 0 to 1, with 0 representing the lowest observed vulnerability and 1 representing highest observed vulnerability. The first reason for a no data value in a raster layer is a result of census data not being provided in an area or being indicated as no population. In the case of water, pixels were masked if determined to be water according to CIESIN's GPWv4: Land and Water Area, Revision 11 (CIESIN, 2018a). An additional layer is provided with a mask for no population according to CIESIN's GPWv4: Population Count, Revision 11 (CIESIN, 2018b).

SVI Category	Coded Value
SVI Value	0 to 1
No data	NA

Data set web page:
SEDAC URL: https://sedac.ciesin.columbia.edu/data/set/usgrid-us-social-vulnerability-index
Permanent URL: https://doi.org/10.7927/6s2a-9r49

Data set format:
The data are available as global grids in GeoTIFF format as a downloadable zip file. Each downloadable is a compressed zip file, which contains: 1) Social Vulnerability Index Grids for the United States for the specified year and projection, and 2) PDF documentation.

Data set downloads:
usgrid-us-social-vulnerability-index-tract-2000-nad83-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2000-wgs84-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2010-nad83-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2010-wgs84-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2014-nad83-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2014-wgs84-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2016-nad83-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2016-wgs84-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2018-nad83-geotiff.zip
usgrid-us-social-vulnerability-index-tract-2018-wgs84-geotiff.zip

IV.　How to Use the Data

The data can be used in any standard Geographic Information System (GIS) and software package for direct mapping and geospatial analysis.

V.　Potential Use Cases

This gridded product enables users to obtain vulnerability metrics along with population estimates for any user-defined area within the U.S., or can be integrated with other spatial data to identify social vulnerability levels in areas such as disaster risk zones, coastal areas, or urban versus rural areas. This can be a significant advantage for understanding patterns of vulnerability in relation to hazard exposure. SEDAC incorporates the data in open map services for use in its own and third party online mapping tools.

VI.　Limitations

The SVI is a relative score based on a locality's relative position with respect to other localities, so changes in SVI scores over time can reflect changes within a locality as well as changes elsewhere. For example, an increase in the SVI score between time periods

does not automatically mean that social vulnerability levels have increased in that locality, but rather it could mean that social vulnerability has declined widely for other parts of the United States. The change of other places to rank higher or lower than the location of interest will change the ranking even if the locality itself stays the same. The CDC produces a version of the SVI in which the tracts and counties within a given state are assigned their 0-1 score relative to other tracts and counties in that state, rather than based on their score relative tracts and counties nation-wide. Users who wish to do state and local level analyses may wish to use the CDC state-by-state SVI data set instead. Likewise, it is not generally recommended to aggregate the SEDAC SVI gridded data to county level as it would be better to use the SVI county product that is scored relative to all counties in the U.S. One advantage of the SEDAC SVI grids is that one can average SVI scores across state and county borders.

The CDC SVI data is based on census data which has its own limitations such as uncertainty in the ACS survey estimates, especially for smaller units such as block groups and census tracts (Spielman et al., 2014, Bazuin and Frazier, 2013). One issue is that the 2018 ACS data lacks socioeconomic data for 238 populated tracts, and therefore a general SVI and socioeconomic scores were not calculated by the CDC for these specific tracts, but three of the sub-category theme SVI scores were calculated. Where there is missing data for a tract at the center of a grid cell, the gridding process draws on the SVI values for any other overlapping tracts. These gaps vary across years, which means that depending on the year, any given pixel may draw on different tracts. This can result in time series discontinuities. In other words, small changes in included tracts and therefore relative rank over years may not be representative of the trend of the country in general or tracts more specifically, but could be owing to differences in the availability of input data in a given year.

Some users may be interested in considering social vulnerability clipped only to areas of population instead of gridded maintaining the polygon-type extent of the original census tract spatial scale. This capability is provided as a secondary layer option labelled "NoPop" by adding an additional mask using CIESIN's GPWv4: Population Count, Revision 11 (CIESIN, 2018b) indicating grid cells with no population. However, it is acknowledged that GPW does not capture population in all areas, especially sporadic rural housing, so it is provided as an optional layer that, as with all population products, should be used with discretion.

These data are available in the original NAD83 projection as well as the WGS84 (Geographic) projection. Small spatial errors may result from the reprojection into WGS84.

VII. Acknowledgments

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(s):

Center for International Earth Science Information Network, (CIESIN), Columbia University. 2021. U.S. Social Vulnerability Index Grids. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/6s2a-9r49. Accessed DAY MONTH YEAR.

XI. Source Code

No source code is provided for this data set.

XII. References

Bazuin, J. T. and J. C. Frazier. 2013. How the ACS gets it wrong: The story of the American Community Survey and a small, inner city neighborhood. *Applied Geography* 45: 292–302.
http://static1.squarespace.com/static/546e4181e4b041f61f02c55c/t/566c0d63df40f3a7317a010e/1449921891566/How+the+ACS+gets+it+wrong+-the+story+of+the+American+Community+Survey+and+a+small,+inner+city+neighborhood_Bazuin_2013.pdf.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018a. Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4Z60M4Z.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018b. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5.

Centers for Disease Control and Prevention (CDC). 2015. Planning for an Emergency: Strategies for Identifying and Engaging At-Risk Groups. A guidance document for Emergency Managers: First edition. Atlanta (GA): CDC. URL: https://www.cdc.gov/nceh/hsb/disaster/atriskguidance.pdf.

Centers for Disease Control and Prevention (CDC)/Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program. CDC Social Vulnerability Index 2000, 2010, 2014, 2016, and 2018. Database US. URL: https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html. Accessed on 25 January 2021.

Cutter, S. L. 2005. The Role of Vulnerability Science in Disaster Preparedness and Response. Research Subcommittee of the US House of Representatives Committee on Science, pp. 1–11.
https://www.start.umd.edu/sites/default/files/files/announcements/CutterCongressionalBriefing.pdf.

Flanagan, B. E., E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, and B. Lewis. 2011. A Social Vulnerability Index for Disaster Management. Journal of Homeland Security and Emergency Management, 8(1).
https://svi.cdc.gov/A%20Social%20Vulnerability%20Index%20for%20Disaster%20Management.pdf.

Flanagan, B. E., E. J. Hallisey, E. Adams, and A. Lavery. 2018. Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers for Disease Control

and Prevention's Social Vulnerability Index. *Journal of Environmental Health*, 80(10), 34-36. PMID: 32327766; PMCID: PMC7179070 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7179070/).

IPCC. 2012. Glossary of terms. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 555-564.

Spielman, S. E., D. Folch, and N. Nagle. 2014. Patterns and causes of uncertainty in the American community survey. *Applied Geography* 46: 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002.

Tate, E. 2013. Uncertainty Analysis for a Social Vulnerability Index. *Annals of the Association of American Geographers*, 103(3), 526-543. https://doi.org/10.1080/00045608.2012.700616.

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	ORCID	Contributors	Revisions
August 24, 2021	0000-0002-8875-4864	Alex de Sherbinin, Carolynne Hultquist	This document is the 1st instance of documentation.

Appendix 3. Data Files

The Social Vulnerability Index Grids zip files contain 1 km rasters in GeoTIFF format fit to the GPW grid for the United States including Alaska and Hawaii for the specified year and projection, as well as the documentation in PDF format. There are 5 categories of layers for each of the 5 years of data with each year having one layer representing overall SVI and 4 sub-themes including Socioeconomic, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation. The composite percentile rankings of census variables for each category at the census tract level are stored as an SVI score, which ranges between 0 and 1, to indicate relatively low or high vulnerability, respectively. An additional projection specification is added to the end of each layer name to indicate either NAD83 or WGS84. All the layers have a mask for water, but "NoPop" is indicated at the end of the filenames for rasters with a mask applied for areas of no population according to GPW.