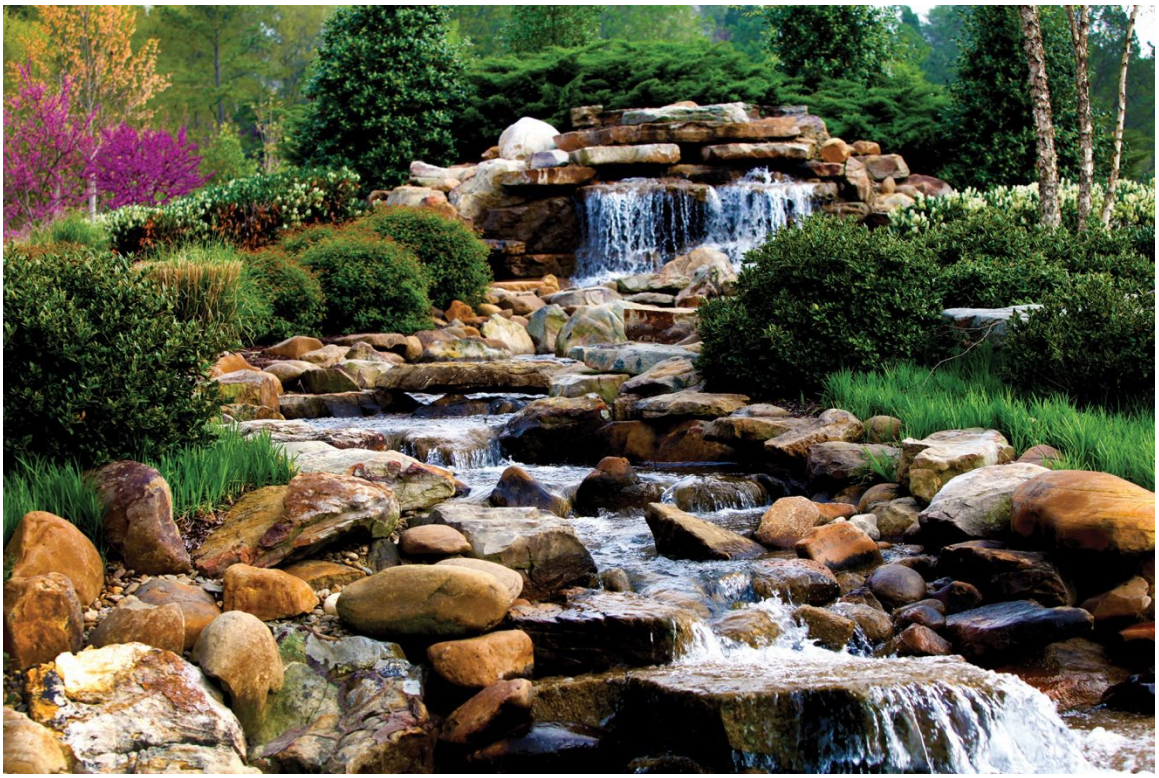
ORNL/SPR-2023/2930
PUB ID 195499

VA EDH Data Curation Documentation FY23-Q3



Hilda B. Klasky
Kevin Sparks
Alina Peluso
Jeremy Logan
Alec Hamaker
Michael McGee
James VanDerslice
Heidi Hanson
Rochelle Watson
Anuj Kapadia

June 2023

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/SPR-2023/2930
PUB ID 195499

Computational Sciences & Engineering Division

REPORT TITLE VA EDH DATA CURATION DOCUMENTATION FY23-Q3

Hilda B. Klasky
Kevin Sparks
Alina Peluso
Jeremy Logan
Alec Hamaker
Michael McGee
James VanDerslice
Heidi Hanson
Rochelle Watson
Anuj Kapadia

June 2023

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

1. INTRODUCTION

The health and well-being of the Nation's men and women who have served in uniform is the highest priority for the U.S. Department of Veterans Affairs (VA). VA is committed to providing timely access to high-quality, recovery-oriented, evidence-based mental health care that anticipates and responds to Veterans' needs and supports the reintegration of returning Service members into their communities. Since its creation, VA has been working to eliminate suicide among all veterans by developing and implementing innovative suicide prevention approaches and resources.

Health outcomes, such as suicide are typically modeled as a function of genetics and environment, where environment refers to factors beyond medical, e.g., air quality, access to transportation and food, homelessness status, etc. Mental health outcomes for each individual are considered to be associated with multiple stressors that fall under a variety of categories including socioeconomic, economic, physical environment. Understanding the relationships between these stressors, covariates and health outcomes requires curated, standardized data that can be input into the VA's Recovery Engagement and Coordination for Health, Veterans Enhanced Treatment (REACH VET) or other health outcomes model.

Environmental Determinants of Health (EDH) as defined by the World Health Organization (WHO) refers to clean air, stable climate, adequate water, sanitation and hygiene, safe use of chemicals, protection from radiation, healthy and safe workplaces, sound agricultural practices, health-supportive cities and built environments, and a preserved nature are all prerequisites for good health.

1.1 BACKGROUND

With funding from the VA Office of Mental Health and Suicide Prevention (OMHSP), the EDH project has developed novel datasets associated with select health outcomes, a methodology for converting spatiotemporal data from one spatial reference (such as a 1km grid) to another (such as US Census Tracts), and health outcomes modeling capabilities. The datasets are an advancement to the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health (SDoH) covariates as key gaps are addressed. We include a finer spatial resolution (Census Tract), and environmental covariates.

The process of curating and standardizing these datasets is non-trivial, as they are often measured at different spatial and temporal resolutions and have different spatial and temporal granularities. For example, the US Census data products typically use census blocks, block groups, or counties, whereas air pollutants from the US Environmental Protection Agency (EPA) and weather data are available on 1km grids, and some economic data may be available only at a zip code level. In this context, standardized refers to the datasets all being at the same scale of spatial extent (e.g., US Census Tract and/or County), and curated refers both to a process that is repeatable, has data provenance, and which uses appropriate methodologies for converting covariates.

The data contained in the EDH datasets are drawn from multiple sources, and variables may have differing degrees of availability, patterns of missing data, and methodological considerations across sources, geographies, and years.

2. DOCUMENTATION OVERVIEW

This data source documentation report provides researchers with information on the structure and contents of the datasets, as well as descriptions of the data sources utilized to populate the data files. This document covers EDH's Fiscal Year 2023, Third Quarter (FY23-Q3) dataset curation documentation. The datasets included are as follows:

1. Educational Attainment 2016-2021, by County.
2. US Population-Weighted Average Elevation 2010, by County.
3. Eviction Rates 2000-2018, by County (update).
4. Food Insecurity 2010-2021, by County.

2.1 RECOMMENDED CITATION FOR THE FY23-Q3 DATASET PACKAGE

Klasky, H.B., Sparks, K., Peluso, A., K., Logan, J., Hamaker, A., McGee, M., VanDerslice, J., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q3, ORNL/SPR-2023/2930 PUB ID 195499, 2023.

2.2 PREVIOUS DOCUMENT RELEASES

Since its beginning, the following EDH project releases (datasets and data curation documentation) have been made available:

For the EDH Data Curation Documentation delivered on FY23-Q2, which contains the following datasets:

- Opioid Prescribing Rate 2013-2020
- Social Capital Atlas Civil Engagement 2022
- Social Capital Atlas Cohesiveness 2022
- Social Capital Atlas Economic Connectedness 2022
- Poverty from the US Census Bureau 2018-2021
- Local Unemployment Rate from the US Bureau of Labor Statistics 2018-2021
- Total Household Income from the US Census Bureau 2018-2021
- Rural-Urban Continuum Codes from the US Census Bureau and US Department of Agriculture 2013.

Refer to: Klasky, H.B., Sparks, K., Peluso, A., Whitehead, M., K., Logan, J., Hamaker, A., McGee, M., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q2, ORNL/SPR-2023/2857, PUB ID 19179. 2023.

For the EDH Data Curation Documentation delivered on FY23-Q1, which contains the following datasets:

- US Area Deprivation Index 2020
- Good Samaritan Laws 2018
- Naloxone Laws 2017

- Opioid Mortality Rate 2014 - 2018
- Opioid Prescribing Rate 2019

refer to: Klasky, H.B., Sparks, K., Logan, J., Hamaker, A., Whitehead, M., Peluso, A., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q1, ORNL/SPR-2022/2694, PUB ID 187842. 2022. https://www.osti.gov/biblio/1909101

For the EDH Data Curation Documentation delivered on FY22-Q4, which contains the following datasets:

- National Mental Health Services Survey
- National Survey on Drug Use and Health
- Occupational Employment and Wage Statistics

refer to: Klasky, H.B., Sparks, K., Logan, J., Hamaker, A., Whitehead, M., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q4, ORNL/SPR-2022/2587, PUB ID 183700. 2022. https://www.osti.gov/biblio/1892396

For the EDH Data Curation Documentation delivered on FY22-Q3, which contains the following datasets:

- Medicare Part D Opioid Prescribing Rates
- High Intensity Drug Trafficking Areas (HIDTA)
- Small-Area Estimates of Housing Characteristics
- Internet Access Services
- Facebook Social Connectedness Index
- Veteran Population for the Civilian Population

refer to: Klasky, H.B., Sparks, K., Logan, J., Tuccillo, J., Whitehead, M., Hamaker, A., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q3, Rev. 2. ORNL/SPR-2022/2487 - Pub ID 178645. 2022. https://www.osti.gov/biblio/1876283

For the EDH Data Curation Documentation delivered on FY22-Q2, which contains the following datasets:

- Eviction Rates (by county)
- Income Inequality (American Community Survey Income Inequality Measures Based on Income to Poverty Ratio by Census Block Group)
- Individual-Oriented Social Vulnerability Index (IOSVI), Census Block Groups, and
- National Instant Criminal Background Check System (NICS), Lethal Means Access.

refer to: Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Rastogi, D., Branstetter, M., Whitehead, M., Hamaker, A., and Watson, R., VA EDH Data Curation Documentation - FY22-Q2, Rev. 2, ORNL/SPR-2022/2391 - Pub ID 174092. 2022. https://www.osti.gov/biblio/1862127-va-edh-data-curation-documentation-fy22-q2

For the EDH Data Curation Documentation delivered on FY22-Q1, which contains the following datasets:

- Social Capital Index Dataset (2019 - updated)
- Social Vulnerability Index Dataset (2018)
- Block Group Area Deprivation Index Dataset for Washington, DC (2019)
- Low Food Access Area Dataset for Washington, DC (2017)

refer to: Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Devineni, P., and Watson, R. VA EDH Data Curation Documentation - FY22-Q1, Rev. 2, ORNL/SPR-2022/2316- Pub ID 172755. 2022. https://www.osti.gov/biblio/1854460-va-edh-data-curation-documentation-fy22-q1

Consult the following source for the FY21 dataset curation documentation: Christian, J.B., Branstetter, M, Klasky, H.B., Tuccillo, J., Sparks, K., Rastogi, D., Watson, R., Yoon, H-J., Kim, Y., VA EDH Data Curation Documentation - FY 2021, Rev. 2, ORNL/SPR-2021/2366 - Pub ID 170648. 2021. https://www.osti.gov/biblio/1854468-va-edh-data-curation-documentation-fy

3. CONTENTS AND STRUCTURE

3.1 DATASET CURATION DOCUMENTATION STANDARD FORMAT

Each data source description follows a standard format with the following fields:

- Sponsor (name of the organization that provided the raw data, e.g., Health Resources and Services Administration [HRSA] for the Area Health Resources Files [AHRF])
- Description (brief, general description of the data)
    - Inclusion in the EDH datasets
        - Lists the social or environmental determinants of health domains to which the data source has contributed variables.
        - Includes additional information about the data source relevant to the EDH dataset.
- Resources (links to original data source documentation, data download sites, and other relevant information).
- Update frequency: how often is each dataset going to be updated.

- Variable definitions and specifications (in tabular format):
  - Variable name
  - Variable label
  - Source table, if multiple data tables were available from the original data source (optional)
  - Numerator (for derived variables, optional)
  - Denominator (for derived variables) or original variable (when renamed for the EDH dataset, optional)
  - total_rows: how many rows are in each column in each dataset (Starting in FY23Q2)
  - null_rows: for each column in each dataset, how many rows are null. (Starting in FY23Q2)
- Variable availability across years (in tabular format):
  - Variable name
  - Data year availability (e.g., 2009 to 2018)

3.2 DATASET CONVENTIONS

Regarding datasets' versioning, Microsoft SQL Server database system is used to supply the datasets. A table is used to hold each dataset. The following schema names for the quarterly releases are (or will be) included in the database: OMHSP_FY22Q1, OMHSP_FY22Q2, OMHSP_FY22Q3, OMHSP_FY22Q4, OMHSP_FY23Q1, …, and so on. These will aid in distinguishing between releases when we deliver the same data set, updated, from one release to another.

Variables in the EDH dataset were created from these several data sources in one of two ways:

1. Drawn directly from the original data source. When the data was available from the data source as needed, we renamed the original variables for clarity and consistency across years, and to fit the naming conventions of the SDOH beta data files.

2. Derived using data from the original data source. For some data sources, it was necessary to calculate percentages or rates for inclusion in the beta data files. The numerators and denominators for the variables and their sources are shown following each data source description.

The following conventions were followed in constructing the EDH datasets to provide researchers with a consistent and easy-to-use resource:

- Variable assignment to annual datasets. Variables appear in the annual datasets that correspond with (1) the single year represented by the original data source (e.g., US Area Deprivation Index 2020), or (2) the last year in a period represented by the data (e.g., American Community Survey data aggregated over 2012 to 2016 is in the 2016 dataset).

- Variable availability. The availability of each variable changes across data years. Following each data source description in this report is a table showing the availability of each variable in the annual datasets. When a variable is not available, we indicate it with NA (not available) or simply '-'.

- Variable naming. Except for the geographic ID variables, all variable names begin with a data source acronym followed by an underscore and a descriptive title.

- Missing values. The datasets use a blank to denote a missing value, almost exclusively. The one exception is the provider ratio variables from the County Health Rankings (CHR) data, which have negative values for counties where the number of providers is zero. This is described further in the description of the CHR data.

Detailed information about each data source is included in the following sections of this report.

3.3 META TABLE

Starting with release FY23Q1, the ORNL team has been supplying an updated metadata table called SEDH_meta_table, which is located in the OMHSP schema. SEDH stands for Social and Environmental Determinants of Health repository. The columns of the SEDH_meta_table are as follows:

- schema. The schema names for the quarterly releases are included in the database: OMHSP_FY22Q4, OMHSP_FY23Q1, OMHSP_FY23Q2,…, and so on;

- table_name. The table name as it appears in the MS SQL Server database;

- table_name_description. The description of the table name;

- column_name. Each of the dataset's column names as they appear in the MS SQL Server table;

- column_name_descriptition. The description of the column name;

- availability_across_years. The years for which the data is available;

- reference_report. The reference to the ORNL's report that has this dataset's data curation documentation;

- report_url. The URL link to the ORNL report;

- column_type. The column type as it appears in the MS SQL Server table;

- column_length. The column length as it appears in the MS SQL Server table.

- total_rows: how many rows are in each column in each dataset. (Starting in FY23Q2)

- null_rows: for each column in each dataset, how many rows are null. (Starting in FY23Q2)

With each new quarterly release, the metadata table will be updated with new details on the aforementioned columns for each delivered dataset.

Please take note that the report_url column with the URL link will be updated in the VA's CDW transmit database as soon as it becomes accessible on the Office of Scientific and Technical Information website (osti.gov) of the US Department of Energy, which is typically four weeks after each quarterly release.

3.4 ERROR CHECKING

Beginning with the FY23Q1 release, the ORNL team will additionally give succinct information regarding error checking activities in order to provide formal evidence that the datasets supplied have been thoroughly error checked. Our data profiling process is described in our manuscript: [1] Klasky, H.B., Hanson, H., Sparks, K., Whitehead, M., Blair, C., and Kapadia, A., "Dataset Repository for Investigating Suicide Risk Using Social and Environmental Determinants of Health", Manuscript under review, Pub ID 183902. 2022.

"Following standard data and software development methodologies, data profiling is performed in four different work environments: 1) a team-shared work environment for selection, extraction, and refinement of raw data (development); 2) an ORNL intranet work environment focused on quality assurance testing (QA-Intra); 3) an ORNL Knowledge Discovery Infrastructure (KDI) secure work environment that stores highly sensitive data and ensures its security (QA-KDI). And finally, 4) a production environment housed within the KDI environment and accessible to our VA sponsors, (Production). We carried out test iterations in each of the four work environments as the datasets moved through them to confirm data integrity and system compatibility.

All datasets were error-checked using a data profiling strategy that includes at least two reviewers and the following test groups:

1. evaluating missingness: i.e. determining the amount of missing data by randomly checking for them;

2. compiling descriptive statistics, such as the number of rows, columns, and types of variable data;

3. appending checksums to a subset of the columns on both the source and destination copies to ensure consistency;

4. consistently representing the social and physical environment using FIPS codes as geographic administrative boundaries and confirming that the FIPS codes correspond to the geographic administrative boundaries of the original data;

5. manually comparing the first, last, and five additional randomly selected rows for consistency between the source and target datasets.

When datasets are developed at ORNL, which we call 'derivative', ORNL will provide extra error-checking utilizing a combination of statistical methodologies based on each dataset's properties, in addition to the data profiling methodology described above."[1]

Descriptive statistics from source datasets were identical for this FY23Q3 delivery when they were imported into our databases.
The error-checking results for FY23Q3 follows:

			Development		QA-Intra		QA-KDI (View)		Production (Transmit)		
Dataset Name	Rows	Columns	Passes	Fails	Passes	Fails	Passes	Fails	Passes	Fails	Error ratio
OMHSP_FY23Q3 - education_county_2016_2021	19322	32	5	0	5	0	5	0	5	0	0
OMHSP_FY23Q3 - elevation_county_2010	3142	3	4	1	5	0	5	0	5	0	0.05
OMHSP_FY23Q3 - eviction_county_2000_2018	59717	8	5	0	5	0	4	1	4	1	0.11
OMHSP_FY23Q3 - food_insecurity_county_2010_2021	37710	19	5	0	5	0	5	0	5	0	0

3.5 FIPS

ORNL employs the Federal Information Processing Standards (FIPS) as geographic identifiers and, in database terms, as the primary key in each dataset or table for the purposes of this project. The FIPS of the United States are a set of publicly announced standards that the National Institute of Standards and Technology (NIST) developed for use in computer systems and non-military applications. Particularly, the US government has developed various FIPS specifications to standardize codes for geographical areas in which FIPS is a unique identifier associated with a region of space or geographical area. The specifications for different geographical areas are as follows: FIPS 10-4 is for country codes or region codes, FIPS 5-2 is for state codes, and FIPS 6-4 is for counties.

Smaller geographic organizations' FIPS codes are generally unique inside bigger geographic entities. FIPS state codes, for example, are unique within a country, while FIPS county codes are unique within a state. Since counties nest inside states, a full county FIPS code specifies both the state and the nesting county. For example, there are 49 counties in the 50 states that finish with the number "001". To make these county FIPS codes distinct, the state FIPS codes are appended to the front of each county (01001, 02001, 04001, etc.), where the first two numbers relate to the state the county is in, and the final three digits correspond to the county explicitly.

The use of FIPS in the US has continued, despite NIST's gradual withdrawal initiated in 2002. NIST intended to replace FIPS with the Geographical Name Information System (GNIS) Feature ID in early 2002. Census.gov defines a geographic identifier, or GEOID, as a "unique identifier to geographic entities that facilitates organization and presentation." GEOIDs are maintained by the US Board on Geographic Names, however after 10 years of replacement efforts, on 2010, and until now, many federal organizations in the United States were still using FIPS. For example, the US Census Bureau employed FIPS codes to identify legal and statistical entities for counties, because neither FIPS codes nor GNIS codes provide appropriate coverage of many smaller geographic areas due to the fuzziness and vagueness of natural boundaries. As a matter of fact, the US Census Bureau creates and maintains codes that include census divisions, regions, tracks, block groups, census blocks, and urban areas. A depiction of the hierarchy is found at the US Census Bureau's standard hierarchy of census geographic entities is found here https://www2.census.gov/geo/pdfs/reference/geodiagram.pdf

As the main key for all the datasets created for this project, we have followed the convention of utilizing the column "FIPS" as the column name to uniquely identify the data, regardless of the source FIPS granularity. We indicate the FIPS granularity, such as region, state, county, census division, tracks, group blocks, etc. in the metadata table and reports' descriptions. We presume that users of these datasets are familiar with joining several datasets using FIPS columns with different geographic area levels.

4. EDUCATION ATTAINMENT BY COUNTY

4.1 SPONSOR

US Census Education Attainment.

4.2 DESCRIPTION

The educational attainment of the US population refers to the highest level of education completed. The educational attainment of the US population is comparable to that of many other industrialized countries, with the vast majority of the population having completed secondary education and a rising number of college graduates surpassing high school dropouts. The US population as a whole spends more years in formal educational programs than other countries. However, income levels vary by race, age, household configuration, and geography.

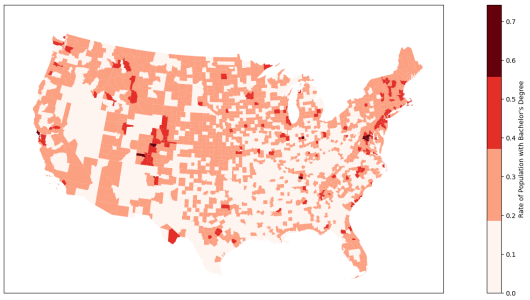


Figure 1. US rate of county population, 18 and up, with a bachelor's degree in 2021

4.3 INCLUSION

Year: 2016-2021 data

Geographical unit: FIPS County Level, Continental US.

Note: It should be noted that the final three rows in the table of column descriptions below were calculated by the ORNL team and are not available to the original data set as found in the US Census Education Attainment.

4.4 RESOURCES

US Census Education Attainment:

Data: Education Attainment

4.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 1 . Education Attainment by County (EDUCATION)

variable name	description
fips	Federal Information Processing Standards (FIPS) - at county level.
year	The year the data pertain to.
pop_18_24_all	Population 18 to 24 years.
nohsd_18_24	Population 18 to 24 years, less than high school graduate.
hsd_18_24	Population 18 to 24 years, high school graduate (includes equivalency).
ad_18_24	Population 18 to 24 years, some college or associate's degree.
bd_18_24	Population 18 to 24 years, bachelor's degree or higher.
pop_25up_all	Population 25 years and over.
ls9_25up	Population 25 years and over, less than 9th grade.
btwn9_12_25up	Population 25 years and over, 9th to 12th grade, no diploma.
hsd_25up	Population 25 years and over, high school graduate (includes equivalency).
scnd_25up	Population 25 years and over, some college, no degree.
ad_25up	Population 25 years and over, associate's degree.
bd_25up	Population 25 years and over, bachelor's degree.
gd_25up	Population 25 years and over, graduate, or professional degree.
hsdup_25up	Population 25 years and over, high school graduate or higher.
bdup_25up	Population 25 years and over, bachelor's degree or higher.
pop_25_34_all	Population 25 to 34 years.
hsdup_25_34	Population 25 to 34 years, high school graduate or higher.
bdup_25_34	Population 25 to 34 years, bachelor's degree or higher.
pop_35_44_all	Population 35 to 44 years.
hsdup_35_44	Population 35 to 44 years, high school graduate or higher.
bdup_35_44	Population 35 to 44 years, bachelor's degree or higher.

pop_45_64_all	Population 45 to 64 years.
hsdup_45_64	Population 45 to 64 years, high school graduate or higher.
bdup_45_64	Population 45 to 64 years, bachelor's degree or higher.
pop_65up_all	Population 65 years and over.
hsdup_65up	Population 65 years and over, high school graduate or higher.
bdup_65up	Population 65 years and over, bachelor's degree or higher.
hsd_rate_18up	(Derivative data) Rate of population, 18 and up, with a high school diploma or equivalent.
bd_rate_18up	(Derivative data) Rate of population, 18 and up, with a bachelor's degree.
gd_rate_18up	(Derivative data) Rate of population, 18 and up, with a graduate degree.

Table 2 . Variable availability across years, (EDUCATION)

variable name	2016	2017	2018	2019	2020	2021	total rows	null rows
fips	X	X	X	X	X	X	19322	0
year	X	X	X	X	X	X	19322	0
pop_18_24_all	X	X	X	X	X	X	19322	0
nohsd_18_24	X	X	X	X	X	X	19322	0
hsd_18_24	X	X	X	X	X	X	19322	0
ad_18_24	X	X	X	X	X	X	19322	0
bd_18_24	X	X	X	X	X	X	19322	0
pop_25up_all	X	X	X	X	X	X	19322	0
ls9_25up	X	X	X	X	X	X	19322	0
btwn9_12_25up	X	X	X	X	X	X	19322	0
hsd_25up	X	X	X	X	X	X	19322	0
scnd_25up	X	X	X	X	X	X	19322	0
ad_25up	X	X	X	X	X	X	19322	0
bd_25up	X	X	X	X	X	X	19322	0
gd_25up	X	X	X	X	X	X	19322	0
hsdup_25up	X	X	X	X	X	X	19322	6440
bdup_25up	X	X	X	X	X	X	19322	6440
pop_25_34_all	X	X	X	X	X	X	19322	0
hsdup_25_34	X	X	X	X	X	X	19322	0
bdup_25_34	X	X	X	X	X	X	19322	0
pop_35_44_all	X	X	X	X	X	X	19322	0
hsdup_35_44	X	X	X	X	X	X	19322	0
bdup_35_44	X	X	X	X	X	X	19322	0

pop_45_64_all	X	X	X	X	X	X	19322	0
hsdup_45_64	X	X	X	X	X	X	19322	0
bdup_45_64	X	X	X	X	X	X	19322	0
pop_65up_all	X	X	X	X	X	X	19322	0
hsdup_65up	X	X	X	X	X	X	19322	0
bdup_65up	X	X	X	X	X	X	19322	0
hsd_rate_18up	X	X	X	X	X	X	19322	0
bd_rate_18up	X	X	X	X	X	X	19322	0
gd_rate_18up	X	X	X	X	X	X	19322	0

5. US POPULATION-WEIGHTED AVERAGE ELEVATION BY COUNTY

5.1 SPONSOR

US Geological Survey.

5.2 DESCRIPTION

A geographic location's elevation is its height above or below a set reference point, most typically a reference geoid, which is a mathematical model of the Earth's sea level as an equipotential gravitational surface. In this dataset, we have compiled individual state files from the U.S. Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA) into one file and included the FIPS code at the county level. In this dataset, elevations are provided in both feet and meters. It should be noted that the average spatial elevation has been adjusted (weighted) for areas where people are typically present.

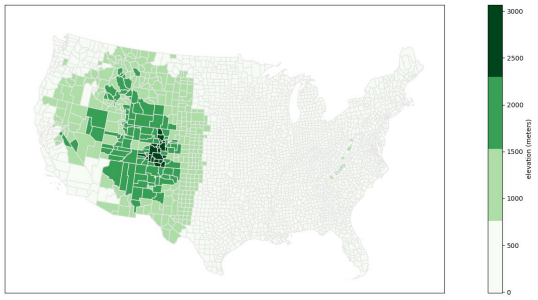


Figure 2. US County-Weighted Population Average Elevation from the 2010 US Geological Survey

5.3 INCLUSION

Year: 2010 data.

Geographical unit: FIPS County Level, Continental US.

5.4 RESOURCES

For more information regarding this dataset or the 2010 US Geological Survey see:

- James A. Vanderslice, PhD
- Danielson, J. J., & Gesch, D. B. (2011). Global multi-resolution terrain elevation data 2010 (GMTED2010) | U.S. Geological Survey.
- World Population Summary- WorldPop. (2020). WorldPop: Population Counts
- US Census Bureau. (2018). Cartographic Boundary Files - Shapefile.

5.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 3 . US Population-Weighted Average Elevation by County (ELEVATION)

variable name	description
fips	Federal Information Processing Standards (FIPS) 5 digit code at county level.
mean_elevation_meters	mean elevation of populated areas in meters.
mean_elevation_feet	mean elevation of populated areas in feet.

Table 4 . Variable availability across years, (ELEVATION)

variable name	2010	total rows	null rows
fips	X	3142	0
mean_elevation_meters	X	3142	0
mean_elevation_feet	X	3142	0

6. EVICTION RATES BY COUNTY

6.1 SPONSOR

Eviction Lab

6.2 DESCRIPTION

This is a delivery update for the eviction dataset from FY22Q2. In the FY22Q2 delivery, eviction rates for United States counties with selected socioeconomic characteristics on evictions from 2000 to 2016 was included. However, the FY22Q2 data is currently accessible on the Eviction Lab website under the "legacy-data" section (Desmond, Matthew, Ashley Gromis, Lavar Edmonds, James Hendrickson, Katie Krywokulski, Lillian Leung, and Adam Porton. "Eviction lab national database: Version 1.0." https://data-downloads.evictionlab.org/#legacy-data).
The Eviction Lab appears to no longer be using that data. The Eviction Lab has changed the data, calculation methods, and variable naming conventions. We are doing all we can to keep the column names as similar as feasible to the updated source. As a result, the reader may notice that the column names in this release differ from those in the dataset presented during FY22Q2. The Eviction Lab obtained court data directly from a number of states. However, several governments either did not consolidate their eviction data or were hesitant to disclose it. Consequently, the Eviction Lab obtained larger datasets of public eviction records from two different companies: LexisNexis Risk Solutions and American Information Research Services Inc. We used the county_propretary_valid_2000_2018.csv file as source for this dataset because of its completeness (it contains more than twice as much data for a given year as court-issued data). These data has been made available by the Eviction Lab under the terms of the Open Data Commons Attribution License (ODC-BY 1.0).

Note: Data for Prince George's County, Maryland, has been omitted from the eviction map to avoid distortion, as the value there is significantly higher than all other values (109.38, compared with the next highest value, which is around 40).

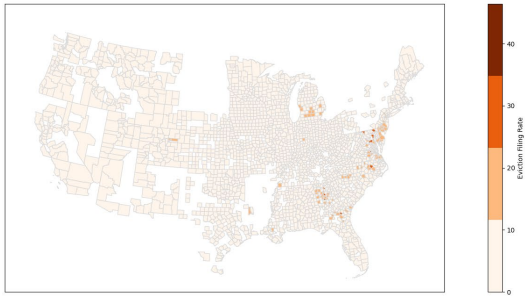


Figure 3. US eviction filings rate per 100 renting households in 2018

6.3 INCLUSION

Year: 2000 - 2018 data

Geographical unit: FIPS County Level, Continental US.

6.4 RESOURCES

- Gromis, Ashley, Ian Fellows, James R. Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond. Estimating Eviction Prevalence across the United States. Princeton University Eviction Lab. https://data-downloads.evictionlab.org/#estimating-eviction-prevalance-across-us/. Deposited May 13, 2022.
- Eviction Lab
- Data for download
- county_proprietary_valid_2000_2018 CSV file data
- Methods and Validation FAQ
- Evictions Supplemental Information
- Open Data Commons Attribution License, ODC-BY 1.0

6.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 5 . Eviction Rates by County (EVICTIONS)

variable name	variable label
fips	Federal Information Processing Standards (FIPS) - at county level.
year	The year the data pertain to.
filings	Number of eviction case filings observed in proprietary data.
filing_rate	Number of eviction case filings per 100 renting households.
threatened	Number of households threatened with eviction observed in proprietary data.
threatened_rate	Number of households threatened with eviction per 100 renting households.
judgements	Number of judgements (households displaced due to eviction) observed in proprietary data.
judgement_rate	Number of judgements per 100 renting households.

Table 6 . Variable availability across years, (EVICTIONS)

variable name	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	total rows	null rows
fips	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	0
year	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	0
filings	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14073
filing_rate	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14086
threatened	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14073
threatened_rate	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14086
judgements	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14073
judgement_rate	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	59717	14086

7. FOOD INSECURITY BY COUNTY

7.1 SPONSOR

Feeding America

7.2 DESCRIPTION

According to the USDA, food insecurity is the inability to consistently obtain enough food for each member of a household to lead an active, healthy life. A family may be in this circumstance for a short while or for a very long time. One indicator of how many people cannot afford food is food insecurity. According to the US Census, approximately 38 million people in America were poor in 2021, including 9 million children. Lower income is the most prevalent cause of food insecurity.

This image presents the rate of population that is food insecure, as defined by Feeding America, in USA counties during 2021.

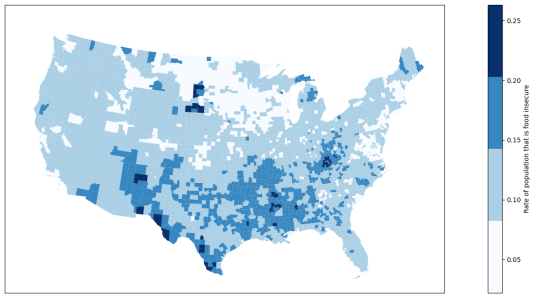


Figure 4. US Food Insecurity Rate in 2021

7.3 INCLUSION

Year: 2010-2021 data.

Geographical unit: FIPS County Level, Continental US.

7.4 RESOURCES

For more information on food insecurity:

- Hake, M., Engelhard, E., & Dewey, A. (2023). Map the Meal Gap 2023: An Analysis of County and Congressional District Food Insecurity and County Food Cost in the United States in 2021. Feeding America.
- Recomended citation for the data analyzed in the report above: Gundersen, C., Strayer, M., Dewey, A., Hake, M., & Engelhard, E. (2023). Map the Meal Gap 2023: An Analysis of County and Congressional District Food Insecurity and County Food Cost in the United States in 2021. Feeding America.
- Feeding America's Map the Meal Gap 2023 Report

- Technical Brief 2023
- Map the Meal Gap Data
- Data Request Form
- Food insecurity

7.5 UPDATE FREQUENCY

Every fiscal year, or as requested by the sponsor, this dataset will be updated and distributed. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

Table 7 . Food Insecurity by County (FOODINSECURITY)

variable name	description
fips	Federal Information Processing Standards (FIPS) - at county level.
year	The year the data pertain to.
food_insr_rate	Percentage of inhabitants who are food insecure.
food_insr_num	Number of inhabitants who are food insecure.
food_insr_blk	Percentage of Black inhabitants (all ethnicities) who are food insecure.
food_insr_his	Percentage of Hispanic inhabitants (all races) who are food insecure.
food_insr_wht	Percentage of white, non-Hispanic inhabitants who are food insecure.
low_income_lmt	Lowest gross income limit as a percent of the federal poverty guidelines for any federal nutrition program.
high_income_lmt	Highest gross income limit as a percent of the federal poverty guidelines for any federal nutrition program.
per_below_low	The percentage of food insecure individuals (FII) who live in households with income at or below low threshold in state.
per_below_high	The percentage of FII who live in households with income above low threshold and below or equal to high threshold in state.
per_above_high	The percentage of FII who live in households with income above high threshold in state.
per_above_low	The percentage of FII who live in households with income above low threshold in state.
chld_insr_rate	Percentage of children (under 18) who are food insecure.
chld_insr_num	Number of children (under 18) who are food insecure.
chld_below_185	Percent of food insecure children who live in households with income below 185% of the federal poverty line.
chld_above_185	Percent of food insecure children who live in households with income above 185% of the federal poverty line.
weekly_cost	The amount of money needed by a food-insecure person to meet weekly food needs.
pop_weekly_cost	Weighted weekly dollars needed by the food insecure multiplied by the population of the geography.

Table 8 . Variable availability across years, (FOODINSECURITY)

variable name	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	total rows	null rows
fips	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
year	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
food_insr_rate	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
food_insr_num	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
food_insr_blk	X	X	X	X	X	X	X	X	X	X	X	X	37710	33204
food_insr_his	X	X	X	X	X	X	X	X	X	X	X	X	37710	32369
food_insr_wht	X	X	X	X	X	X	X	X	X	X	X	X	37710	28439
low_income_lmt	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
high_income_lmt	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
per_below_low	X	X	X	X	X	X	X	X	X	X	X	X	37710	0
per below high	X	X	X	X	X	X	X	X	X	X	X	X	37710	10103
per_above_high	X	X	X	X	X	X	X	X	X	X	X	X	37710	3143
per_above_low	X	X	X	X	X	X	X	X	X	X	X	X	37710	28282
chld_insr_rate	X	X	X	X	X	X	X	X	X	X	X	X	37710	7
chld insr num	X	X	X	X	X	X	X	X	X	X	X	X	37710	30
chld_below_185	X	X	X	X	X	X	X	X	X	X	X	X	37710	9
chld_above_185	X	X	X	X	X	X	X	X	X	X	X	X	37710	9
weekly_cost	X	X	X	X	X	X	X	X	X	X	X	X	37710	3
pop_weekly_cost	X	X	X	X	X	X	X	X	X	X	X	X	37710	3