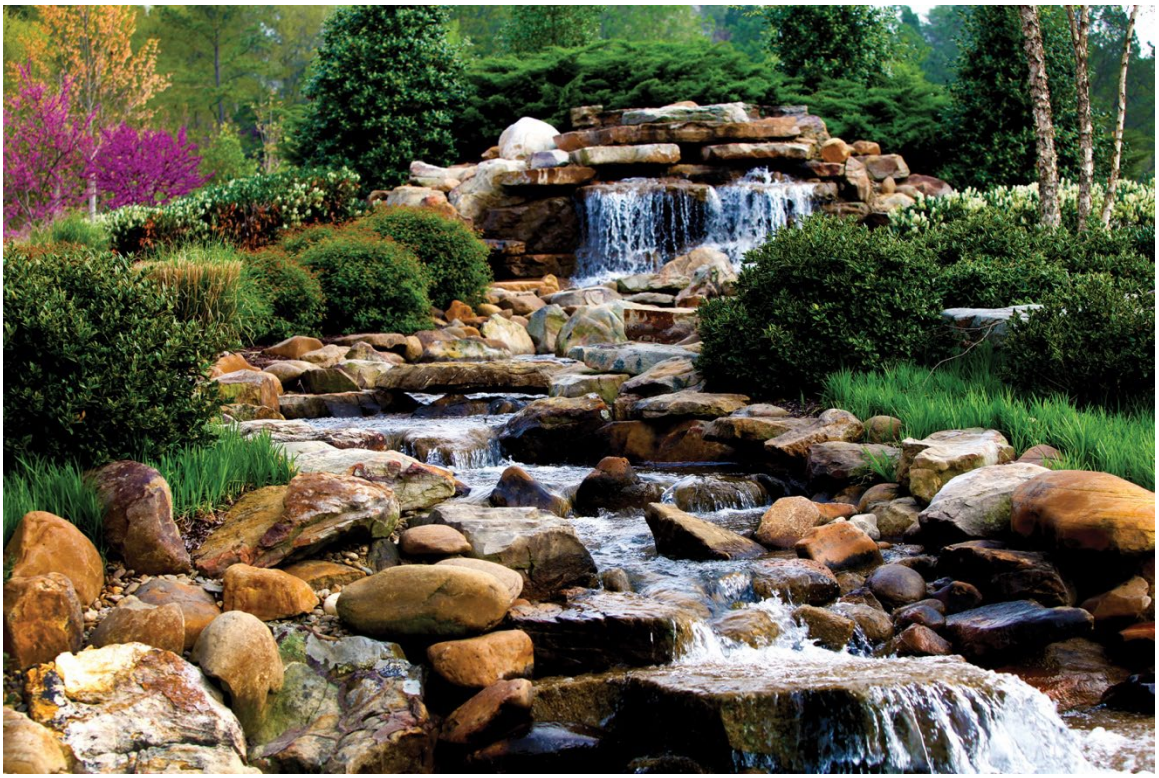
ORNL/SPR-2022/2487
Pub ID 178645

VA EDH Data Curation Documentation FY22-Q3, Rev. 2



Hilda B. Klasky
Kevin Sparks
Jeremy Logan
Joe Tuccillo
Matthew Whitehead
Alec Hamaker
Heidi Hanson
Rochelle Watson
Anuj Kapadia

June 2022

OAK RIDGE National Laboratory

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

DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via OSTI.GOV.

Website www.osti.gov

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone 703-605-6000 (1-800-553-6847)
TDD 703-487-4639
Fax 703-605-6900
E-mail info@ntis.gov
Website http://classic.ntis.gov/

Reports are available to US Department of Energy (DOE) employees, DOE contractors, Energy Technology Data Exchange representatives, and International Nuclear Information System representatives from the following source:

Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831
Telephone 865-576-8401
Fax 865-576-5728
E-mail reports@osti.gov
Website https://www.osti.gov/

ORNL/SPR-2022/2487
Pub ID 178645

Computational Sciences & Engineering Division

VA EDH DATA CURATION DOCUMENTATION FY22-Q3, REV. 2

Hilda B. Klasky
Kevin Sparks
Jeremy Logan
Joe Tuccillo
Matthew Whitehead
Alec Hamaker
Heidi Hanson
Rochelle Watson
Anuj Kapadia

June 2022

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

CHANGE HISTORY

Rev No.	Change	Performed by:	Date
2	1. Fixed table 10 variable availability across years	Hilda B. Klasky	Nov, 2022

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. VA is 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 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) is 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, a finer spatial resolution (Census Tract), and environmental covariates are included.

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 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 dataset 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 FY22-Q3 dataset curation documentation.

2.1 RECOMMENDED CITATION FOR FY22-Q3 DATASET PACKAGE

Klasky, H.B., Sparks, K., Logan, 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.

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 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.

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.

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.

3. CONTENTS AND STRUCTURE

The data in the EDH datasets were gathered from a variety of publicly available data sources. In FY22-Q3 release the data sources included follow:

- Medicare Part D Opioid Prescribing Rates (2019).
- High Intensity Drug Trafficking Areas (HIDTA) (2018 - 2021)
- Small-Area Estimates of Housing Characteristics (2019)
- Internet Access Services (2019)
- Facebook Social Connectedness Index (2021)
- Veteran Population Status for the Civilian Population (2020)

3.1 DATASET CURATION DOCUMENTATION STANDARD FORMAT

Each data source description follows a standard format with the next 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 SDOH 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)
- Variable availability across years (in tabular format)
  - Variable name
  - Data year availability (e.g. 2009 to 2018)

3.2 DATASET CONVENTIONS

Regarding the datasets' versioning, Microsoft SQL Server database system is used to store the datasets. 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 from one release to the next.

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., Nursing Home Compare data for facilities in 2016 appears in the 2016 county dataset), 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.
- 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.

4. MEDICARE PART D OPIOID PRESCRIBING RATES

4.1 SPONSOR

Centers for Medicare & Medicaid Services (CMS)

4.2 DESCRIPTION

The Medicare Part D Opioid Prescribing Rates by Geography dataset is based on information collected from CMS administrative claims data for Medicare Part D enrollees accessible from the CMS Chronic Condition Data Warehouse. The information is derived entirely from final-action Part D prescription medication claims.

4.3 INCLUSION

The Medicare Part D Opioid Prescribing Rates by Geography dataset includes FIPS geographic comparisons of the quantity and proportion of Medicare Part D opioid prescriptions. The CMS publishes this data in yearly updates; this is the 2019 version of this data. This dataset's columns that aren't immediately obvious are described in the table below.

Column Name	Description
Tot_Prscrbrs	The number of providers prescribing Medicare Part D drugs.
Tot_Opioid_Prscrbrs	The number of Medicare Part D providers prescribing opioid drugs.
Tot_Opioid_Clms	The number of Medicare Part D opioid drug claims, including original prescriptions and refills.
Tot_Clms	The number of Medicare Part D drug claims, including original prescriptions and refills.
Opioid_Prscrbng_Rate	The number of opioid related drug claims divided by the overall drug claims and multiplied by 100.
Opioid_Prscrbng_Rate_5Y_Chg	The percentage point difference in the opioid prescribing rate from five years previous to the data year, which is calculated by subtracting the rate five years previous from the rate in the data year. At each geographic level, the change in the prescribing rate is displayed as an increase, decrease, or no change. At the state level, an increase reflects a percentage point difference of at least 0.10 and a decrease reflects a difference of at least -0.10; at the county -level, an increase reflects a difference of at least 1.0 and a decrease reflects a difference of at least -1.0.
Opioid_Prscrbng_Rate_1Y_Chg	The percentage point difference in the opioid prescribing rate from one year previous to the data year, which is calculated by subtracting the rate one year previous from the rate in the data year. At each geographic level, the change in the prescribing rate is displayed as an increase, decrease, or no change. At the state level, an increase reflects a percentage point difference of at least 0.10 and a decrease reflects a difference of at least -0.10; at the county -level, an increase reflects a difference of at least 1.0 and a decrease reflects a difference of at least -1.0.

LA_Tot_Opioid_Clms	The number of Medicare Part D opioid drug claims that are considered long-acting, including original prescriptions and refills.
LA_Opioid_Prscrbng_Rate	The number of Long-Acting opioid drug claims divided by the opioid drug claims and multiplied by 100.
LA_Opioid_Prscrbng_Rate_5Y_Chg	The percentage point difference in the long-acting opioid prescribing rate from five years previous to the data year, which is calculated by subtracting the rate five years previous from the rate in the data year. At each geographic level, the change in the prescribing rate is displayed as an increase, decrease, or no change. At the state level, an increase reflects a percentage point difference of at least 0.10 and a decrease reflects a difference of at least -0.10; at the county -level, an increase reflects a difference of at least 1.0 and a decrease reflects a difference of at least -1.0.
LA_Opioid_Prscrbng_Rate_1Y_Chg	The percentage point difference in the long-acting opioid prescribing rate from one year previous to the data year, which is calculated by subtracting the rate one year previous from the rate in the data year. At each geographic level, the change in the prescribing rate is displayed as an increase, decrease, or no change. At the state level, an increase reflects a percentage point difference of at least 0.10 and a decrease reflects a difference of at least -0.10; at the county -level, an increase reflects a difference of at least 1.0 and a decrease reflects a difference of at least -1.0.

4.4 RESOURCES

For more information on the Medicare Part D Opioid Prescribing Rates:

Main: https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-medicaid-opioid-prescribing-rates/medicare-part-d-opioid-prescribing-rates-by-geography

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. By request, we can provide data going back to 2013.

Table 1. Medicare Part D Opioid Prescribing Rates (HEALTHCARE_2019)

variable name	variable label
FIPS	Federal Information Processing Standards (FIPS), county level fips codes.
State	The name of the state.
County	The name of the county.
Tot_Prscrbrs	Part D Prescribers.
Tot_Opioid_Prscrbrs	Part D Opioid Prescribers.
Tot_Opioid_Clms	Opioid Claims.
Tot_Clms	Overall Claims.
Opioid_Prscrbng_Rate	Opioid Prescribing Rate.
Opioid_Prscrbng_Rate_5Y_Chg	Five Year Change in Opioid Prescribing Rate.
Opioid_Prscrbng_Rate_1Y_Chg	One Year Change in Opioid Prescribing Rate.
LA_Tot_Opioid_Clms	Long-Acting Opioid Claims.
LA_Opioid_Prscrbng_Rate	Long-Acting Opioid Prescribing Rate.
LA_Opioid_Prscrbng_Rate_5Y_Chg	Five Year Change in Long-Acting Opioid Prescribing Rate
LA_Opioid_Prscrbng_Rate_1Y_Chg	One Year Change in Long-Acting Opioid Prescribing Rate

Table 2. Variable availability across years, (HEALTHCARE_2019)

variable name	2019
FIPS	X
State	X
County	X
Tot_Prscrbrs	X
Tot_Opioid_Prscrbrs	X
Tot_Opioid_Clms	X
Tot_Clms	X
Opioid_Prscrbng_Rate	X
Opioid_Prscrbng_Rate_5Y_Chg	X
Opioid_Prscrbng_Rate_1Y_Chg	X
LA_Tot_Opioid_Clms	X
LA_Opioid_Prscrbng_Rate	X
LA_Opioid_Prscrbng_Rate_5Y_Chg	X
LA_Opioid_Prscrbng_Rate_1Y_Chg	X

5. HIGH INTENSITY DRUG TRAFFICKING AREAS (HIDTA)

5.1 SPONSOR

Washington/Baltimore High Intensity Drug Trafficking Areas Program

5.2 DESCRIPTION

The High Intensity Drug Trafficking Areas (HIDTA) program, established by Congress with the Anti-Drug Abuse Act of 1988, offers support to federal, state, local, and tribal law enforcement agencies working in locations identified as important drug-trafficking areas in the United States.

5.3 INCLUSION

The data was obtained from the High Intensity Drug Trafficking Areas Program's activities from 2018 to 2021.

5.4 RESOURCES

For more information about HIDTA:

Main: https://www.dea.gov/operations/hidtas

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. High Intensity Drug Trafficking Areas (HIDTA)

variable name	variable label
FIPS	Federal Information Processing Standards (FIPS), county level fips codes.
state	US state name.
county	The county name.
seizure_date	The date of seizure.
drug	The type of drug seized.
quantity	The quantity of drugs seized.
unit	The weight unit of measurement (kilogram - Kg, or Deci atomic mass Unit = D.U.)

Table 4. Variable availability across years (HIDTA)

variable name	2018	2019	2020	2021
FIPS	X	X	X	X
state	X	X	X	X
county	X	X	X	X
seizure_date	X	X	X	X
drug	X	X	X	X
quantity	X	X	X	X
unit	X	X	X	X

6. SMALL-AREA ESTIMATES OF HOUSING CHARACTERISTICS

6.1 SPONSOR

United States Census Bureau

6.2 DESCRIPTION

Small-area estimates of selected housing characteristics at the census block level for 2019. Topics include year of construction, residential type and density, rent burden, homeownership costs, and property values.

6.3 INCLUSION

Data was acquired from the American Community Survey (ACS) 2019 5-Year Estimates for the District of Columbia, Maryland, Pennsylvania, and Virginia.

Year of home construction and residential type/density were acquired directly from the ACS Summary File (SF). Rent burden, homeownership costs, and property values were estimated on a national residential synthetic population produced from the ACS Public-Use Microdata Sample (PUMS) using Oak Ridge National Laboratory's UrbanPop framework. UrbanPop simulates multiple (n = 30) residential synthetic populations, to which the housing attributes of interest were attached via the Census Microdata API. Finally, Monte Carlo estimation was used to produce a single estimate and standard error value for each block group. This dataset columns' source tables are listed in the table below.

Column Name	Source Table
ybl_2014L	SF B25034: Year Structure Built
ybl_2010_2013	SF B25034: Year Structure Built
ybl_2000_2009	SF B25034: Year Structure Built
ybl_1990_1999	SF B25034: Year Structure Built
ybl_1980_1989	SF B25034: Year Structure Built
ybl_1970_1979	SF B25034: Year Structure Built
ybl_1960_1969	SF B25034: Year Structure Built
ybl_1950_1959	SF B25034: Year Structure Built
ybl_1940_1949	SF B25034: Year Structure Built
ybl_1939E	SF B25034: Year Structure Built
dwg_01u_d	SF B25024: Units in Structure
dwg_01u_a	SF B25024: Units in Structure
dwg_02u	SF B25024: Units in Structure
dwg_03_04u	SF B25024: Units in Structure
dwg_05_09u	SF B25024: Units in Structure
dwg_10_19u	SF B25024: Units in Structure
dwg_20_49u	SF B25024: Units in Structure
dwg_GE50u	SF B25024: Units in Structure
dwg_moho	SF B25024: Units in Structure

dwg_informal	SF B25024: Units in Structure
avg_rbr	PUMS GRPIP: Gross rent as a percentage of income past 12 months
avg_rbr_se	PUMS GRPIP: Gross rent as a percentage of income past 12 months
rbr_GE30	PUMS GRPIP: Gross rent as a percentage of income past 12 months
rbr_GE30_se	PUMS GRPIP: Gross rent as a percentage of income past 12 months
avg_ocs	PUMS OCPIP: Selected monthly owner costs as a percentage of household income during the past 12 months
avg_ocs_se	PUMS OCPIP: Selected monthly owner costs as a percentage of household income during the past 12 months
ocs_GE30	PUMS GRPIP: Gross rent as a percentage of income past 12 months
ocs_GE30_se	PUMS GRPIP: Gross rent as a percentage of income past 12 months
avg_pvl	PUMS VALP: Property value
avg_pvl_se	PUMS VALP: Property value

6.4 RESOURCES

For more information on the ACS:

Main: https://www.census.gov/programs-surveys/acs

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. Small-Area Estimates of Housing Characteristics (HOUSING_2019)

variable name	variable label
FIPS	US Census Block Group Federal Information Processing Standards (FIPS) code.
ybl_2014L	Proportion of total housing units in dwellings built 2014 or later
ybl_2010_2013	Proportion of total housing units in dwellings built 2010 - 2013
ybl_2000_2009	Proportion of total housing units in dwellings built 2000 - 2009
ybl_1990_1999	Proportion of total housing units in dwellings built 1990 - 1999
ybl_1980_1989	Proportion of total housing units in dwellings built 1980 - 1989
ybl_1970_1979	Proportion of total housing units in dwellings built 1970 - 1979
ybl_1960_1969	Proportion of total housing units in dwellings built 1960_1969
ybl_1950_1959	Proportion of total housing units in dwellings built 1950_1959
ybl_1940_1949	Proportion of total housing units in dwellings built 1940_1949
ybl_1939E	Proportion of total housing units in dwellings built 1939 or earlier
dwg_01u_d	Proportion of total housing units in single-family detached dwellings
dwg_01u_a	Proportion of total housing units in single-family attached dwellings
dwg_02u	Proportion of total housing units in 2-unit dwellings
dwg_03_04u	Proportion of total housing units in 3 - 4 unit dwellings
dwg_05_09u	Proportion of total housing units in 5 - 9 unit dwellings
dwg_10_19u	Proportion of total housing units in 10 - 19 unit dwellings
dwg_20_49u	Proportion of total housing units in 20 - 49 unit dwellings
dwg_GE50u	Proportion of total housing units in dwellings with 50 units or more
dwg_moho	Proportion of total housing units in mobile homes
dwg_informal	Proportion of total housing units in informal housing (Boat, RV, van, etc.)
avg_rbr	Average rent burden for based on gross rent
avg_rbr_se	Monte Carlo standard error of average rent burden for based on gross rent costs
rbr_GE30	Proportion households with a rent burden 30% of income or greater based on gross rent costs
rbr_GE30_se	Monte Carlo standard error of proportion households with a rent burden 30% of income or greater based on gross rent costs
avg_ocs	Average selected monthly owner costs (mortgage, insurance, etc.)
avg_ocs_se	Monte Carlo standard error of average selected monthly owner costs (mortgage, insurance, etc.)
ocs_GE30	Proportion households with homeownership costs 30% of household income or greater
ocs_GE30_se	Monte Carlo standard error of proportion households with homeownership costs 30% of household income or greater
avg_pvl	Average property value of owner-occupied housing
avg_pvl_se	Monte Carlo standard error of average property value of owner-occupied housing

Table 6. Variable availability across years, (HOUSING_2019)

variable name	2019
FIPS	X
ybl_2014L	X
ybl_2010_2013	X
ybl_2000_2009	X
ybl_1990_1999	X
ybl_1980_1989	X
ybl_1970_1979	X
ybl_1960_1969	X
ybl_1950_1959	X
ybl_1940_1949	X
ybl_1939E	X
dwg_01u_d	X
dwg_01u_a	X
dwg_02u	X
dwg_03_04u	X
dwg_05_09u	X
dwg_10_19u	X
dwg_20_49u	X
dwg_GE50u	X
dwg_moho	X
dwg_informal	X
avg_rbr	X
avg_rbr_se	X
rbr_GE30	X
rbr_GE30_se	X
avg_ocs	X
avg_ocs_se	X
ocs_GE30	X
ocs_GE30_se	X
avg_pvl	X
avg_pvl_se	X

7. INTERNET ACCESS SERVICES

7.1 SPONSOR

Federal Communications Commission

7.2 DESCRIPTION

Census tract level Internet Access Services' status as of June 30, 2019. This dataset provides data from residential fixed connections per 1000 households by census tract (N=73,767).

7.3 INCLUSION

The description of the code per connections follows:

Code	Connections per 1,000 Households
0	Zero
1	Zero < x <= 200
2	200 < x <=400
3	400 < x <=600
4	600 < x <=800
5	800 < x

7.4 RESOURCES

For more information on Internet Access Services:
Main: https://www.fcc.gov/internet-access-services-reports

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. Internet Access Services (INTERNET_ACCESS)

variable name	variable label
FIPS	Federal Information Processing Standards (FIPS) at census tract level.
pcat_all	Residential Fixed High-Speed Connections over 200 kbps in at least one direction per 1000 Households
pcat_10x1	Residential Fixed High-Speed Connections at least 10 Mbps downstream and at least 1 Mbps upstream per 1000 Households

Table 8. Variable availability across years, (INTERNET_ACCESS)

variable name	2019
FIPS	X
pcat_all	X
pcat_10x1	X

8. FACEBOOK SOCIAL CONNECTEDNESS INDEX

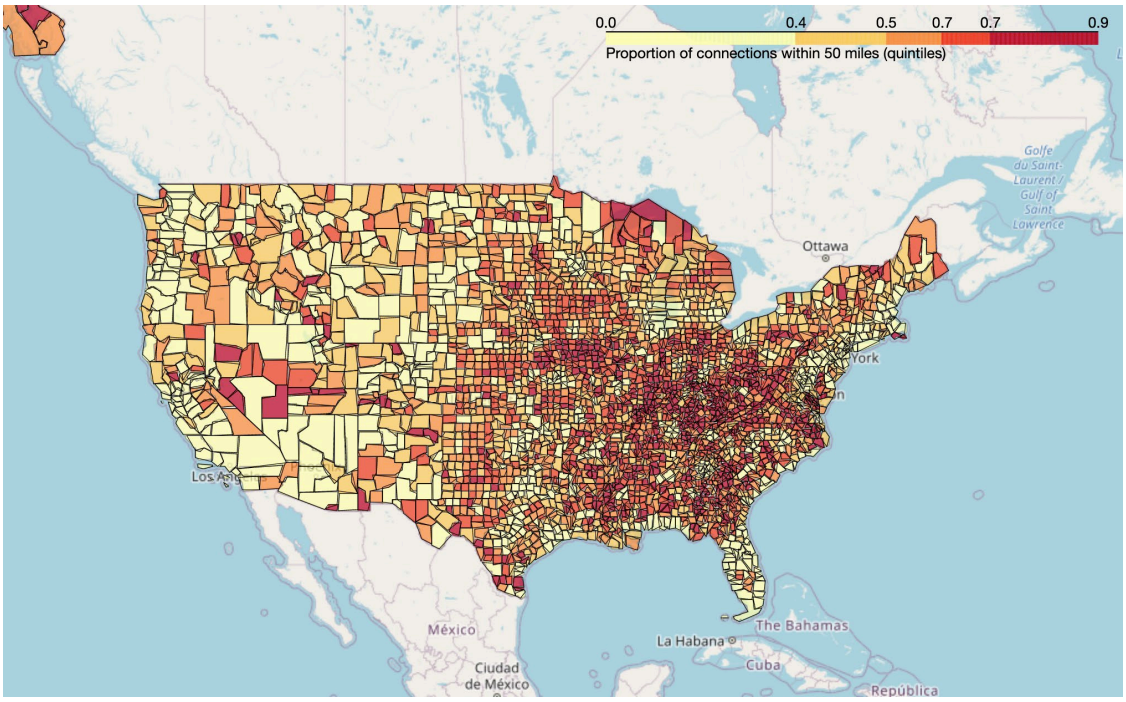
8.1 SPONSOR

Facebook

8.2 DESCRIPTION

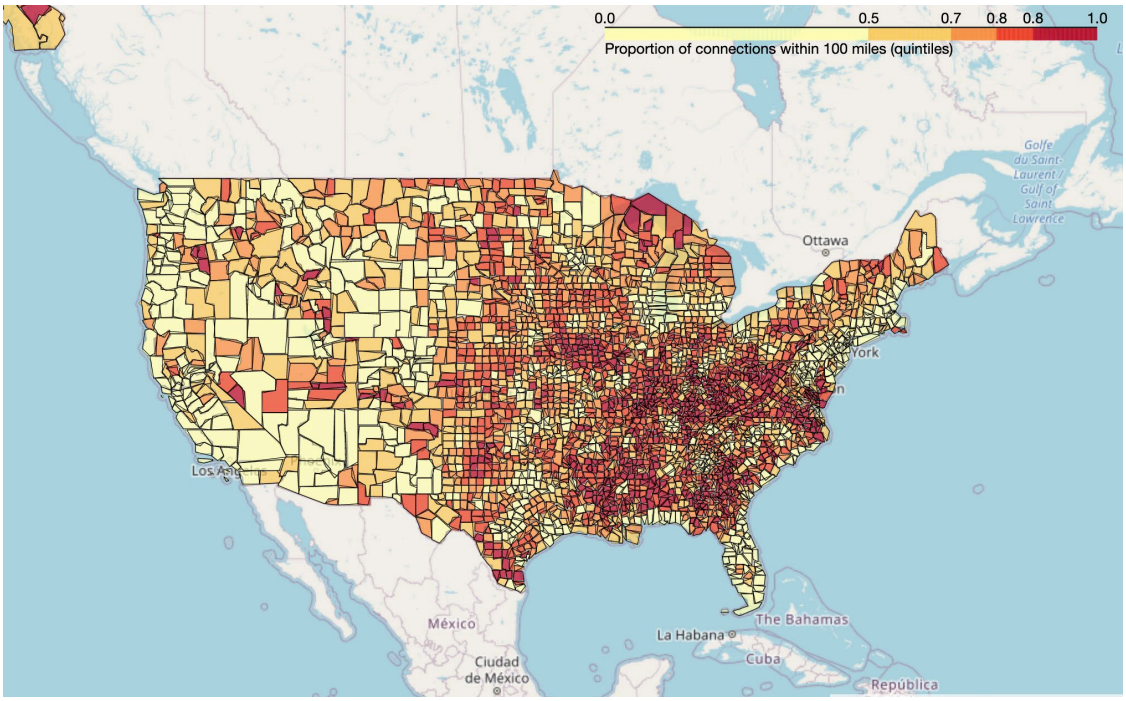
The Social Connectedness Index (SCI) is a measure of social connectivity across regions. This index assesses the degree to which two geographic regions are related as reflected by Facebook friendship relationships. It specifically assesses the likelihood that two people in two different places are Facebook friends with each other. SCI can provide valuable insights on economics, social mobility, and health. Thus, we utilized Facebook's SCI and data to compute the likelihood of one person being related with someone else within 50, 100, and 500 miles, which is the subject of this dataset. In the images below, yellow represents a lesser possibility of social connectedness, while red suggests a larger likelihood.

The image below presents the Facebook data map with proportion of connections within 50 miles.

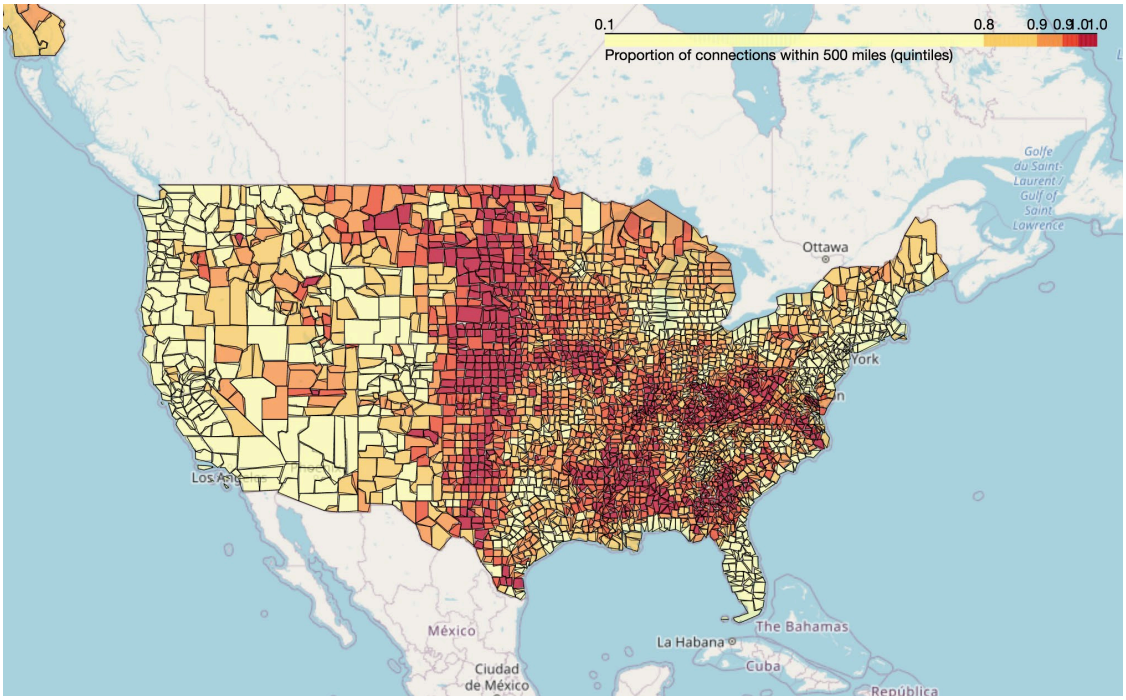


Social Connectedness Within 50 miles map

The images below depict the Facebook data maps with proportion of connections within 100 and 500 miles.



Social Connectedness Within 100 miles map



Social Connectedness Within 500 miles map

For more information, see reference Bailey, Michael, et al., Social connectedness in urban areas, Journal of Urban Economics 118 (2020), 103264.

8.3 INCLUSION

We investigated the geographical distribution of social networks in the United States using anonymized and aggregated data from Facebook. To protect user anonymity, we only include FIPS in our research that

have a total population of at least 500 individuals. These social network statistics are combined with information from the 2015 Census Bureau 5-year American Community Survey (ACS) and the 2014 Internal Revenue Service (IRS) Individual Income Tax Statistics.

In addition, we provide two more measures of connectedness, along with an associated rank for each. The first assesses within-county connectedness, focusing only on connections that fall within the same county. The second measures total connectedness of each county regardless of the location of those connections. Both measures are given on an integer scale ranging from 1 to 1 billion, following the scale established by the original Facebook data. For each of these measures, a rank is also provided that indicates where this county falls when counties are ordered by the measure.

8.4 RESOURCES

For more information on the Facebook Social Connectedness Index:

Meta Social Connectedness Index

Social connectedness dataset

County location data source

8.5 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 9. Facebook Social Connectedness Index (SOC_CON)

variable name	variable label
FIPS	Federal Information Processing Standards (FIPS), county level fips codes.
within_county_connectedness	Assesses within-county connectedness, focusing only on connections that fall within the same county.
within_county_connectedness_rank	Indicates where this county falls when counties are ordered by the measure.
overall_county_connectedness	Measures total connectedness of each county regardless of the location of those connections.
overall_county_connectedness_rank	Indicates where this county falls when counties are ordered by the measure.
share_50mi	Share of friends that live within 50 miles.
share_100mi	Share of friends that live within 100 miles.
share_500mi	Share of friends that live within 500 miles.

Table 10. Variable availability across years, (SOC_CON)

variable name	**2021**
FIPS	X
within_county_connectedness	X
within_county_connectedness_rank	X
overall_county_connectedness	X
overall_county_connectedness_rank	X
share_50mi	X
share_100mi	X
share_500mi	X

9. VETERAN POPULATION STATUS FOR THE CIVILIAN POPULATION

9.1 SPONSOR

Census Reporter and American Community Survey (ACS).

9.2 DESCRIPTION

Veteran Population Status for the Civilian Population 18 years and over, at county level. As the data includes only population 18 years and older, hence the decrease in overall expected.

9.3 INCLUSION

ACS 2020, 5-year release.

9.4 RESOURCES

For more information on the Veteran Status for the Civilian Population 18 years and over:

https://censusreporter.org/data/table/?table=B21001&geo_ids=01000US,050|01000US&primary_geo_id=01000US#

9.5 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 11. Veteran Population Status for the Civilian Population (VET_POP)

variable name	**variable label**
FIPS	Federal Information Processing Standards (FIPS) at county level.
total_pop	Total Population
total_pop_error	Total Population Error
Veteran_pop	Veteran Population
Veteran_pop_error	Veteran Population Error
Nonveteran_pop	Non-veteran Population
Nonveteran_pop_error	Non-veteran Population Error

Table 12. Variable availability across years, (VET_POP)

variable name	2020
FIPS	X
total_pop	X
total_pop_error	X
Veteran_pop	X
Veteran_pop_error	X
Nonveteran_pop	X
Nonveteran_pop_error	X