

ScHARe



The ScHARe Data Ecosystem Last updated: November 27, 2023

This document is intended to provide a comprehensive list of the datasets available in the ScHARe Data Ecosystem for analysis in the ScHARe Terra instance. Using the ScHARe Data Ecosystem, researchers are able to search, link, share, and contribute to a collection of datasets relevant to social science, health outcomes, minority health and health disparities research.

The collection is comprised of:

- **Google-hosted Public Datasets** - Publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program. Examples: US Census Data; American Community Survey (ACS)
- **ScHARe-hosted Public Datasets** - Publicly accessible, de-identified datasets hosted by ScHARe. Examples: Social Vulnerability Index (SVI), Behavioral Risk Factor Surveillance System (BRFSS)
- **ScHARe-hosted Project Datasets** - Publicly accessible and controlled-access, funded program/project datasets shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy. Example: Jackson Heart Study (JHS)

A detailed list of the datasets available in the ScHARe Data Ecosystem, including links to documentation and other helpful resources for each dataset, is available in the sections below. The datasets are categorized as follows, based on their content:

A - SOCIAL DETERMINANTS OF HEALTH

- **A1 Multiple Categories:** Datasets that include data on multiple Social Determinants of Health (SDoH) factors/indicators
- **A2 Economic Stability:** Datasets that include data on unemployment, poverty, housing stability, food insecurity and hunger, work related injuries, etc.
- **A3 Education Access and Quality:** Datasets that include data on graduation rates, school proficiency, early childhood education programs, interventions to address developmental delays, etc.
- **A4 Health Care Access and Quality:** Datasets that include data on health literacy, use of health IT, emergency room waiting times, evidence-based preventive healthcare, health screenings, treatment of substance use disorders, family planning services, access to a primary care provider and high quality care, access to telehealth and electronic exchange of health information, access to health insurance, adequate oral care, adequate prenatal care, STD prevention measures, etc.
- **A5 Neighborhood and Built Environment:** Datasets that include data on access to broadband internet, access to safe water supplies, toxic pollutants and environmental risks, air quality, blood lead levels, deaths from motor vehicle crashes, asthma and COPD cases and hospitalizations, noise exposure, smoking, mass transit use, etc.
- **A6 Social and Community Context:** Datasets that include data on crime rates, imprisonment, resilience to stress, experiences of racism and discrimination, etc. For incidence and prevalence of anxiety, depression, and other mental health conditions, see section “B1 – Diseases and conditions” below
- **A7 Health Behaviors:** Datasets that include data on health behaviors

B - HEALTH OUTCOMES

- **B1 Diseases and Conditions:** Datasets that include data on incidence and prevalence of specific diseases and health conditions
- **B2 Clinical Data/EHRs:** Datasets that include data from electronic health records

Datasets available in the ScHARe Data Ecosystem

A - SOCIAL DETERMINANTS OF HEALTH

A1 - Multiple Categories

Datasets that include data on multiple Social Determinants of Health (SDoH) factors/indicators.

American Community Survey (ACS): Detailed US demographic data at various geographic resolutions

Source: U.S. Census Bureau

Description:

The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people by contacting over 3.5 million households across the country. The resulting data provides incredibly detailed demographic information across the US aggregated at various geographic levels which helps determine how more than \$675 billion in federal and state funding are distributed each year.

Businesses use ACS data to inform strategic decision-making. ACS data can be used as a component of market research, provide information about concentrations of potential employees with a specific education or occupation, and which communities could be good places to build offices or facilities. For example, someone scouting a new location for an assisted-living center might look for an area with a large proportion of seniors and a large proportion of people employed in nursing occupations.

Through the ACS, we know more about jobs and occupations, educational attainment, veterans, whether people own or rent their homes, and other topics. Public officials, planners, and entrepreneurs use this information to assess the past and plan the future. For more information, see the Census Bureau ACS Information Guide.

Links: <https://console.cloud.google.com/marketplace/product/united-states-census-bureau/acs>

US Census Data: 2000 and 2010 US Census data

Source: U.S. Census Bureau

Description:

The United States census count (also known as the Decennial Census of Population and Housing) is a count of every resident of the US. The census occurs every 10 years and is conducted by the United States Census Bureau. Census data is publicly available through the census website, but much of the data is available in summarized data and graphs. The raw data is often difficult to obtain, is typically divided by region, and it must be processed and combined to provide information about the nation as a whole.

The United States census dataset includes nationwide population counts from the 2000 and 2010 censuses. Data is broken out by gender, age and location using zip code tabular areas (ZCTAs) and GEOIDs. ZCTAs are generalized representations of zip codes, and often, though not always, are the same as the zip code for an area. GEOIDs are numeric codes that uniquely identify all administrative, legal, and statistical geographic areas for which the Census Bureau tabulates data. GEOIDs are useful for correlating census data with other censuses and surveys.

Links: <https://console.cloud.google.com/marketplace/product/unity-us-census-bureau/acs>

Area Deprivation Index (ADI): *An index of socioeconomic status for communities*

Source: BroadStreet

Description:

The Area Deprivation Index (ADI) can show where areas of deprivation and affluence exist within a community. The ADI is calculated with 17 indicators from the American Community Survey (ACS) having been well-studied in the peer-reviewed literature since 2003, and used for 20 years by the Health Resources and Services Administration (HRSA). High levels of deprivation have been linked to health outcomes such as 30-day hospital readmission rates, cardiovascular disease deaths, cervical cancer incidence, cancer deaths, and all-cause mortality. The 17 indicators from the ADI encompass income, education, employment, and housing conditions at the Census Block Group level.

The ADI is available on BigQuery for release years 2018-2020 and is reported as a percentile that is 0-100% with 50% indicating a "middle of the nation" percentile. Data is provided at the county, ZIP, and Census Block Group levels. Neighborhood and racial disparities occur when some neighborhoods have high ADI scores and others have low scores. A low ADI score indicates affluence or prosperity. A high ADI score is indicative of high levels of deprivation. Raw ADI scores and additional statistics and dataviz can be seen in this ADI story with a BroadStreet free account.

Links: <https://console.cloud.google.com/marketplace/product/broadstreet-public-data/adi>

Social Determinants of Health Database: *SDOH variables from 5 key domains: social context, economic context, education, physical infrastructure, and healthcare context*

Source: Agency for Healthcare Research and Quality (AHRQ)

Description:

AHRQ's database on Social Determinants of Health (SDOH) was created under a project funded by the Patient Centered Outcomes Research (PCOR) Trust Fund. The purpose of this project is to create easy to use, easily linkable SDOH-focused data to use in PCOR research, inform approaches to address emerging health issues, and ultimately contribute to improved health outcomes. The database was developed to make it easier to find a range of well documented, readily linkable SDOH variables across domains without having to access multiple source files, facilitating SDOH research and analysis. Variables in the files correspond to five key SDOH domains: social context (e.g., age, race/ethnicity, veteran status), economic context (e.g., income, unemployment rate), education, physical infrastructure (e.g, housing, crime, transportation), and healthcare

context (e.g., health insurance). The files can be linked to other data by geography (county, ZIP Code, and census tract). The database includes data files and codebooks by year at three levels of geography, as well as a documentation file.

Links: <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html>

Community Resilience Estimates: *Community resilience estimates calculated by modeling individual and household characteristics*

Source: U.S. Census Bureau

Description:

Community resilience is the capacity of individuals and households to absorb, endure, and recover from the health, social, and economic impacts of a disaster such as a hurricane or pandemic. Estimates at the tract and county level are calculated by modeling individual and household characteristics, including poverty, crowding, and unemployment, from the 2019 ACS.

Links: <https://www.census.gov/programs-surveys/community-resilience-estimates.html>

Minority Health Social Vulnerability Index: *Indicators that identify communities that may need support before, during, and after disasters*

Source: Geospatial Research, Analysis, and Services Program in the Agency for Toxic Substances and Disease Registry and HHS Office of Minority Health

Description:

In 2021, the Office of Minority Health (OMH) partnered with CDC to launch the Minority Health SVI. The Minority Health SVI is an extension of the CDC SVI that combines the 15 social factors included in the original CDC SVI with additional factors known to be associated with COVID-19 outcomes. The factors are developed using five-year estimates of demographic data from the U.S. Census Bureau's American Community Survey (ACS), and are organized into six themes: • Socioeconomic Status • Household Composition and Disability • Minority Status and Language • Housing Type and Transportation • Health Care Infrastructure and Access • Medical Vulnerability The Minority Health SVI also expands the Minority Status and Language theme to include statistics for specific race and ethnicity categories and languages. Specifically, data is included for the five racial and ethnic minority groups currently defined by the 1997 OMB Directive 15 and the top five languages spoken by populations with limited English proficient (LEP) at the county level.

Links: <https://www.census.gov/programs-surveys/community-resilience-estimates.html>

The Social Vulnerability Index: *Community resilience data to natural or human-caused disasters and disease outbreaks*

Source: Centers for Disease Control

Description:

The Social Vulnerability Index uses U.S. census variables at tract level to help local officials identify communities that may need support in preparing for hazards or recovering from disaster. Social vulnerability refers to the resilience of communities when confronted by external stresses on human health, stresses such

as natural or human-caused disasters, or disease outbreaks. Reducing social vulnerability can decrease both human suffering and economic loss.

Links: <https://svi.cdc.gov/>

500 Cities: Local Data for Better Health: *City- and census-tract-level estimates for chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 US cities*

Source: Centers for Disease Control, Robert Wood Johnson Foundation, and the Centers for Disease Control Foundation

Description:

The 500 Cities Project provides city- and census-tract-level small area estimates for chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 cities in the United States. These small area estimates will allow cities and local health departments to better understand the burden and geographic distribution of health-related variables in their jurisdictions and assist them in planning public health interventions. Included among prevention measures is health insurance status.

Links: <https://chronicdata.cdc.gov/browse?category=500+Cities+%26+Places>

Native Hawaiian and Pacific Islander NHIS: *Health of Native Hawaiians and Pacific Islanders in all 50 states*

Source: Centers for Disease Control

Description:

In 2014, NCHS conducted the Native Hawaiian and Pacific Islander National Health Interview Survey (NHPI NHIS). The NHPI NHIS was a survey in which about 3,000 households containing one or more NHPI residents were surveyed by NHIS field staff using the 2014 NHIS instrument. The NHPI NHIS was an unprecedented opportunity to collect rich and accurate information from a large NHPI sample about the health of Native Hawaiians and Pacific Islanders in all 50 states.

Links: <https://www.cdc.gov/nchs/nhis/nhpi.html>

A2 - Economic Stability

Datasets that include data on unemployment, poverty, housing stability, food insecurity and hunger, work related injuries, etc.

GDP and Income by County: *U.S. income and GDP at the county level*

Source: Bureau of Economic Analysis

Description:

This public dataset was created by the Bureau of Economic Analysis (BEA). It provides a county level view of income, wages, proprietors' income, dividends, interest, rents, and government benefits, including a number of federal and state-level subsidies.

Per capita income can be used to gauge the average financial health and associated social needs of an area. Analysis across regions offers a way to assess relative standard of living and quality of life of the population. Trends analysis of these data over time can also uncover specific regions of economic growth or decline across a variety of indicators.

These personal income data represent an important lens into the financial security and socioeconomic determinants of health at the community level. They are used by the federal government to allocate hundreds of billions of dollars into state and local programs, to project budgets and trust fund balances, and to develop a more complete picture of labor costs. Personal income statistics can also help illustrate the dynamics between Americans' incomes, spending, and savings.

The data summarize per capita income at the county level, including personal income, net earnings, transfer receipts, benefits programs, unemployment insurance, subsidy programs, retirement, dividends, insurance compensation, and several other economic indicators measured by the Department of Commerce or reported to other public agencies. For more information, please refer to the BEA's Regional Economic Accounts Definitions.

Links: <https://console.cloud.google.com/marketplace/product/bureau-economic-analysis/gdp-income-by-county>

US Inflation and Unemployment: *U.S. economic statistics for inflation, prices and unemployment*

Source: U.S. Bureau of Labor Statistics

Description:

This dataset includes economic statistics on inflation, prices, unemployment, and pay & benefits provided by the Bureau of Labor Statistics (BLS).

Links: <https://console.cloud.google.com/marketplace/product/bls-public-data/cpi-unemployment>

Quarterly Census of Employment and Wages: *A quarterly count of employment and wages*

Source: U.S. Bureau of Labor Statistics

Description:

The Quarterly Census of Employment and Wages (QCEW) program publishes a quarterly count of employment and wages reported by employers covering more than 95 percent of U.S. jobs, available at the county, MSA, state and national levels by industry.

The dataset, hosted as part of the Cloud Public Datasets Program , gives county-level information on jobs and wages each quarter starting in 1990. The counties are identified by geoid which can easily be joined with both all FIPS codes or US county boundaries to unlock new insights within the data. Both of these datasets are available in BigQuery through the Cloud Public Datasets.

Links: <https://console.cloud.google.com/marketplace/product/bls-public-data/qcew>

Point-in-Time Homelessness Count: *Annual Homeless Assessment Report to Congress*

Source: U.S. Dept of Housing and Urban Development

Description:

This database contains the data reported in the Annual Homeless Assessment Report to Congress (AHAR). It represents a point-in-time count (PIT) of homeless individuals, as well as a housing inventory count (HIC) conducted annually.

The data represent the most comprehensive national-level assessment of homelessness in America, including PIT and HIC estimates of homelessness, as well as estimates of chronically homeless persons, homeless veterans, and homeless children and youth.

These data can be trended over time and correlated with other metrics of housing availability and affordability, in order to better understand the particular type of housing resources that may be needed from a social determinants of health perspective.

HUD captures these data annually through the Continuum of Care (CoC) program. CoC-level reporting data have been crosswalked to county levels for purposes of analysis of this dataset.

Links: <https://console.cloud.google.com/marketplace/product/housing-urban-development/homelessness-count>

Low Income Housing Tax Credit Program: *Data on low-income housing development across the U.S.*

Source: U.S. Department of Housing and Urban Development

Description:

The Low-Income Housing Tax Credit (LIHTC) program gives State and local agencies the equivalent of nearly \$8 billion in annual budget authority to issue tax credits for the acquisition, rehabilitation, or new construction of rental housing targeted to lower-income households.

The LIHTC database, created by HUD and available to the public since 1997, contains information on over 47,000 projects and 3 million housing units placed in service between 1987 and 2017. It is the only complete national source of information on the size, unit mix, and location of individual projects.

These data have also been geocoded, enabling researchers to look at the geographical distribution and neighborhood characteristics of tax credit projects. It may also help show how incentives to locate projects in low-income areas and other underserved markets are working.

The database includes project address, number of units and low-income units, number of bedrooms, year the credit was allocated, year the project was placed in service, whether the project was new construction or rehab, type of credit provided, and other sources of project financing. For more information, see HUD.gov.

Links: <https://console.cloud.google.com/marketplace/product/housing-urban-development/lihtc-program>

US Residential Real Estate Data: *Comprehensive data and analytics for 100 million U.S. homes*

Source: House Canary

Description:

Instant access to reliable property, loan and valuation information for 100 million homes. Machine learning algorithms process hundreds of data sources to provide Home Price Indices for 381 Metros, 18,300 ZIP codes and 4 million blocks covering >95% of the US residential market. Make investment decisions based on 40-year historical volatility information and 3 year forecasts.

Examples of the more than 70 data points are available: Valuations - Property value, Value forecast, Rental value Property Data - Details, Sales history, Rental report Loan Data - LTV, Mortgage lien, Notice of default

Links: <https://console.cloud.google.com/marketplace/product/housecanary-com/housecanary>

Current Population Survey (CPS) Annual Social and Economic Supplement: *Labor force statistics - annual work activity, income, health insurance, and health*

Source: U.S. Census Bureau; U.S. Bureau of Labor Statistics (BLS)

Description:

CPS, sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS) and fielded by the U.S. Census Bureau. The CPS is the primary source of labor force statistics for the population of the United States. The CPS is the source of numerous high-profile economic statistics, including the national unemployment rate, and provides data on a wide range of issues relating to employment and earnings. The CPS also collects extensive demographic data that complement and enhance our understanding of labor market conditions in the nation. The CPS Annual Social and Economic Supplement is an annual survey sponsored by the U.S. Census Bureau that includes additional detail on annual work activity, income, health insurance, and health.

Links: <http://www.census.gov/cps/about/>

Food Access Research Atlas: *Food access indicators for low-income and other census tracts*

Source: U.S. Department of Agriculture

Description:

The Food Access Research Atlas presents an overview of food access indicators for low-income and other census tracts using different measures of supermarket accessibility; provides food access data for populations within census tracts; and offers census-tract-level data on food access that can be downloaded for community planning or research purposes.

Links: <https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/>

Current Population Survey Food Security Supplement (CPS-FSS): *Food security, food expenditures, and use of food and nutrition assistance programs*

Source: U.S. Department of Agriculture (USDA); U.S. Census; Bureau of Labor Statistics

Description:

The Current Population Survey Food Security Supplement (CPS-FSS) is the source of national and state-level statistics on food insecurity used in USDA's annual reports on household food security. The CPS is a monthly labor force survey of about 50,000 households conducted by the Census Bureau for the Bureau of Labor

Statistics. Once each year, after answering the labor force questions, the same households are asked a series of questions (the Food Security Supplement) about food security, food expenditures, and use of food and nutrition assistance programs. Food security data have been collected by the CPS-FSS each year since 1995.

Links: https://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpssupps

National Health and Nutrition Examination Survey: *Health and nutritional status of adults and children*

Source: Centers for Disease Control

Description:

NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. NHANES is a major program of NCHS. NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Links: <https://wwwn.cdc.gov/nchs/nhanes/Default.aspx>

American Housing Survey (AHS): *Size, composition, and quality of U.S. housing and household characteristics*

Source: U.S. Census Bureau

Description:

The American Housing Survey (AHS) is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The survey has been the most comprehensive national housing survey in the United States since its inception in 1973, providing current information on the size, composition, and quality of the nation's housing and measuring changes in our housing stock as it ages. The AHS is a longitudinal housing unit survey conducted biennially in odd-numbered years, with samples redrawn in 1985 and 2015. The survey provides up-to-date information about the quality and cost of housing in the United States and major metropolitan areas. The survey also includes questions about: the physical condition of homes and neighborhoods; the costs of financing and maintaining homes; and the characteristics of people who live in these homes.

Links: <https://www.census.gov/programs-surveys/ahs/about/upcoming-releases/2021-release.html>

Supplemental Nutrition Assistance Program (SNAP) Policy Database: *State policy options in the Supplemental Nutrition Assistance Program*

Source: U.S. Department of Agriculture

Description:

The SNAP Policy Database provides a central data source for information on State policy options in the Supplemental Nutrition Assistance Program (SNAP). The database draws on policy information from a wide

variety of sources, including surveys by USDA's Food and Nutrition Service (FNS), national and State policy research organizations, State policy manuals, and news articles.

Links: <https://www.ers.usda.gov/data-products/snap-policy-database.aspx>

Supplemental Nutrition Assistance Program (SNAP) Program Data: *Supplemental Nutrition Assistance Program data monthly aggregate counts at state level*

Source: U.S. Department of Agriculture

Description:

The SNAP Program Data has monthly aggregate counts at state level of participating individuals, households, and benefit issuance.

Links: <https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap>

A3 - Education Access and Quality

Datasets that include data on graduation rates, school proficiency, early childhood education programs, interventions to address developmental delays, etc.

U.S. Department of Education ED Facts Data Files: *Adjusted cohort graduation rates and assessment participation and proficiency*

Source: U.S. Department of Education

Description:

ED Facts is a U.S. Department of Education (ED) initiative to collect, analyze, and promote the use of high-quality, pre-kindergarten through grade 12 data. It collects data on district and school demographics, program participation, and performance data and centralizes data provided by state education agencies (SEAs) with other data assets, such as financial grant information, within the Department to enable better analysis and use in policy development, planning and management.

Links: <https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html#acgr>

National Household Education Surveys Program (NHES): *Educational activities*

Source: U.S. Department of Education

Description:

The National Household Education Surveys Program (NHES), the flagship household survey program of the National Center for Education Statistics, collects nationally representative, descriptive data on the educational activities of children and families in the United States. The NHES surveys cover learning at all ages, from early childhood to school age through adulthood. NHES surveys are conducted every three to four years, on various topics. The current NHES topics are early childhood care and education, family involvement in schools, and homeschooling. These data may be used to analyze trends over time.

Links: <https://nces.ed.gov/nhes/>

IPEDS: *Graduation Rates and Financial Aid*

Source: National Center for Education Statistics

Description:

Integrated Postsecondary Education Data System (IPEDS) Gathers information from every college, university, and technical/vocational institution that participates in the federal student financial aid programs (Title IV-eligible institutions). Required by the Higher Education Act of 1965. Describes basic characteristics of institutions, enrollments, completions and completers, graduation rates and other outcome measures, faculty and staff, finances, institutional prices, student financial aid, admissions, and academic libraries Many organizations and individuals rely on IPEDS data, including Congress, Federal and state agencies, schools, educators, media, and private industry.

Links: <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?gotoReportId=7&fromIped=7>

HERD: *Higher education R&D expenditures reported by all institutions*

Source: NSF

Description:

The HERD Survey is the primary source of information on research and development expenditures at U.S. colleges and universities. The survey collects information on R&D expenditures by field of research and source of funds and also gathers information on types of research, expenses, and headcounts of R&D personnel. The survey is an annual census of institutions that expended at least \$150,000 in separately accounted for R&D in the fiscal year.

Links: <https://nces.nsf.gov/surveys/higher-education-research-development/2021#survey-info>

Early Childhood Longitudinal Study (ECLS): *Child assessments and child questionnaires*

Source: NCES

Description:

The ECLS-K:2011 kindergarten–fifth grade public-use data file (NCES 2019-050) provides data at the child level for each of the 18,174 children who participated, or whose parent participated, in at least one of the two kindergarten data collections (Fall 2010 or Spring 2011). Each child record contains data from the child assessments and child questionnaires, the various respondents associated with the child, weights and imputation flags, and administrative variables. This is the final public-use file released for the ECLS-K:2011 and includes data for all nine rounds of data collection (fall 2010, spring 2011, fall 2011, spring 2012, fall 2012, spring 2013, spring 2014, spring 2015, and spring 2016).

Links: <https://nces.ed.gov/pubsearch/getpubcats.asp?sid=024>

A4 - Health Care Access and Quality

Datasets that include data on health literacy, use of health IT, emergency room waiting times, evidence-based preventive healthcare, health screenings, treatment of substance use disorders, family planning services, access to a primary care provider and high quality care, access to telehealth and electronic exchange of health

information, access to health insurance, adequate oral care, adequate prenatal care, STD prevention measures, etc.

Center for Medicare and Medicaid Services - Dual Enrollment: *Dual enrollment of Medicare and Medicaid by county*

Source: U.S. Department of Health & Human Services

Description:

This public dataset was created by the Centers for Medicare & Medicaid Services. The data summarize counts of enrollees who are dually eligible for both Medicare and Medicaid program, including those in Medicare Savings Programs.

“Duals” represent 20 percent of all Medicare beneficiaries, yet they account for 34 percent of all spending by the program, according to the Commonwealth Fund. As a representation of this high-needs, high-cost population, these data offer a view of regions ripe for more intensive care coordination that can address complex social and clinical needs.

In addition to the high cost savings opportunity to deliver upstream clinical interventions, this population represents the county-by-county volume of patients who are eligible for both state level (Medicaid) and federal level (Medicare) reimbursements and potential funding streams to address unmet social needs across various programs, waivers, and other projects.

The dataset includes eligibility type and enrollment by quarter, at both the state and county level. These data represent monthly snapshots submitted by states to the CMS, which are inherently lower than ever-enrolled counts (which include persons enrolled at any time during a calendar year).

Links: <https://console.cloud.google.com/marketplace/product/hhs/dual-enrollment>

Medicare: *Centers for Medicare & Medicaid Services - Medicare utilization*

Source: U.S. Department of Health & Human Services

Description:

This public dataset was created by the Centers for Medicare & Medicaid Services. The data summarizes the utilization and payments for procedures, services, and prescription drugs provided to Medicare beneficiaries by specific inpatient and outpatient hospitals, physicians, and other suppliers. The dataset includes the following data: common inpatient and outpatient services, all physician and other supplier procedures and services, and all Part D prescriptions.

Providers determine what they will charge for items, services, and procedures provided to patients and these charges are the amount that providers bill for an item, service, or procedure.

Links: <https://console.cloud.google.com/marketplace/product/hhs/medicare>

Health Professional Shortage Areas: *Federally designated areas with a shortage of healthcare workers*

Source: U.S. Department of Health & Human Services

Description:

Health Professional Shortage Areas (HPSAs) are federal designations that indicate health care provider shortages. HRSA's Bureau of Health Workforce (BHW) develops shortage designation criteria and uses them to decide whether or not a geographic area or population group is a Health Professional Shortage Area (HPSA), Medically Underserved Area (MUA), or Medically Underserved Population (MUP).

These designated areas indicate where there is a shortage of healthcare professionals, indicating potential limited access to medical services for a geography or a particular population. These data include the degree of shortage (e.g. ratio of providers), geography of the shortage area, and FTE shortage of practitioners. Such information could be used to identify geographies for new programs, partnerships, campaigns, define opportunities to target alternative resources, where to set up clinics, or where telehealth interventions could be particularly valuable.

More than 34 federal programs depend on HPSA/MUA/MUP designations to determine eligibility or as a funding preference. About 20 percent of the U.S. population resides in primary medical care HPSAs. HPSAs may have shortages of primary medical care, dental, or mental health providers; may be urban or rural areas; population groups; or medical or other public facilities.

Links: <https://console.cloud.google.com/marketplace/product/hhs/health-professional-shortage-areas>

Medical Expenditure Panel Survey (MEPS): *Cost and use of healthcare and health insurance coverage*

Source: Agency for Healthcare Research and Quality (AHRQ)

Description:

MEPS is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of healthcare and health insurance coverage. MEPS collects data on the specific health services that Americans use, how frequently they use them, the cost of these services, and how they are paid for, as well as data on the cost, scope, and breadth of health insurance held by and available to U.S. workers.

Links: <https://meps.ahrq.gov/mepsweb/>

HEDIS: *Primary Care Access and Quality Measures*

Source: Dartmouth College

Description:

Data include measures of primary care utilization, quality of care for diabetes, mammography, leg amputation and preventable hospitalizations.

Links: <https://data.dartmouthatlas.org/primary-care/#longitudinal>

CAHPS - Clinician and Group Survey: *Primary Care Access and Quality Measures*

Source: Agency for Healthcare Research and Quality

Description:

The CAHPS Clinician and Group Survey (CG-CAHPS) asks patients about their experiences with providers and staff in primary care and specialty care settings.

Links: https://datatools.ahrq.gov/cahps/?type=tab&tab=cahpscarhps&_gl=1%2A1dztmjo%2A_ga%2AMTMzNjgzNzY2Mi4xNjgyMDg5MjE2%2A_ga_1NPT56LE7J%2AMTY4OTI1NzkwMi4zLjEuMTY4OTI1Nzk0Mi4yMC4wLjA

Medicare Current Beneficiary Survey (MCBS): *Preliminary Evaluation of Public Data Sources on Health Care Quality*

Source: Centers for Medicare & Medicaid Services

Description:

The Medicare Current Beneficiary Survey (MCBS), sponsored by the Centers for Medicare & Medicaid Services Office of Enterprise Data and Analytics (OEDA) through a contract with NORC at the University of Chicago is a continuous, in-person, longitudinal survey of a representative national sample of the Medicare population. It has been carried out continuously for more than 30 years, encompassing more than one million total interviews. The MCBS is designed to aid CMS in administering, monitoring, and evaluating Medicare programs, is the leading source of information on Medicare and its impact on beneficiaries, provides important information on Medicare beneficiaries that is NOT available in CMS administrative data and plays an essential role in monitoring and examining health care access, utilization, and care transition and coordination.

Links: <https://www.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/mcbs-public-use-file>

A5 - Neighborhood and Built Environment

Datasets that include data on access to broadband internet, access to safe water supplies, toxic pollutants and environmental risks, air quality, blood lead levels, deaths from motor vehicle crashes, asthma and COPD cases and hospitalizations, noise exposure, smoking, mass transit use, etc.

National Environmental Public Health Tracking Network: *Environmental indicators and health, exposure, and hazard data*

Source: Centers for Disease Control

Description:

The Tracking Network is a system of integrated health, exposure, and hazard information and data from a variety of national, state, and city sources. Maps, tables, and charts with data about environmental indicators are available, including data on asthma, biomonitoring and population exposures, birth defects, cancer including childhood cancers, childhood lead poisoning, chronic obstructive pulmonary disease, community characteristics, community design, COVID-19, developmental disabilities, drinking water, drought, environmental justice, heart disease and stroke, heat and heat-related illness, hormone disorders, lifestyle risk factors, pesticide exposures, populations and vulnerabilities, precipitation and flooding, radon, reproductive and birth outcomes, sunlight and ultraviolet, toxic substance releases, unintentional carbon monoxide poisoning.

Links: <https://ephtracking.cdc.gov/showHome.action4.8MB4>

Local Area Transportation Characteristics for Households (LATCH): *Average local area transportation data for households at census tract level*

Source: U.S. Department of Transportation (DOT)

Description:

LATCH data provides average weekday household person-miles traveled, person trips, vehicle-miles traveled and vehicle trips at census tract level.

Links: <https://www.bts.gov/latch/latch-data>

Occupational Requirements Survey (ORS): *Work-related environmental conditions and physical, education, training, and experience, as well as mental and cognitive requirements*

Source: Bureau of Labor Statistics

Description:

ORS is an establishment-based survey conducted by BLS. The ORS provides comprehensive data on specific work-related requirements. BLS publishes information on physical demands (e.g., standing, lifting, climbing), environmental conditions (e.g., extreme temperatures, heavy vibrations, hazardous contaminants), education, training, and experience, as well as mental and cognitive requirements (e.g., decision making, adaptability, contacts) for detailed occupations in the U.S. economy.

Links: <https://www.bls.gov/opub/hom/ors/home.htm>

Greenhouse Gas Reporting Program (GHGRP): *Greenhouse Gas Reporting Program*

Source: U.S. Environmental Protection Agency

Description:

The Greenhouse Gas Reporting Program (GHGRP) collects Greenhouse Gas (GHG) data from large emitting facilities, suppliers of fossil fuels and industrial gases that result in GHG emissions when used, and facilities that inject carbon dioxide underground.

Links: <https://www.epa.gov/ghgreporting/data-sets>

A6 - Social and Community Context

Datasets that include data on crime rates, imprisonment, resilience to stress, experiences of racism and discrimination, etc. For incidence and prevalence of anxiety, depression, and other mental health conditions, see section “B1 – Diseases and conditions” below.

Racial Attitudes in America Survey Dataset: *Racial attitudes and experiences of discrimination*

Source: Pew Research Center

Description:

National survey by Pew Research Center conducted Feb. 29-May 8, 2016, among 3,769 adults (including 1,799 whites, 1,004 blacks and 654 Hispanics). The survey is centered primarily around the divide between blacks and whites and on the treatment of black people in the U.S. today.

Links: <https://www.pewresearch.org/social-trends/dataset/2016-racial-attitudes-in-america-survey/>

FBI Hate Crime Statistics: *State level hate crime statistics*

Source: Federal Bureau of Investigation

Description:

The Uniform Crime Reporting (UCR) Program generates reliable statistics for use in law enforcement. It also provides information for students of criminal justice, researchers, the media, and the public. The program has been providing crime statistics since 1930. The UCR Program includes data from more than 18,000 city, university and college, county, state, tribal, and federal law enforcement agencies. Agencies participate voluntarily and submit their crime data either through a state UCR program or directly to the FBI's UCR Program.

Links: <https://ucr.fbi.gov/hate-crime>

County Health Rankings National Findings Report: *Data to support community-led efforts to improve health equity*

Source: University of Wisconsin Population Health Institute

Description:

County Health Rankings & Roadmaps (CHR&R) brings actionable data, evidence, guidance and stories to support community-led efforts to grow community power and improve health equity. The University of Wisconsin Population Health Institute (UWPHI) created CHR&R for communities across the nation, with funding from the Robert Wood Johnson Foundation.

Links: <https://www.countyhealthrankings.org/>

General Social Survey (GSS): *Social characteristics and attitudes*

Source: NORC at the University of Chicago

Description:

For five decades, the General Social Survey (GSS) has studied the growing complexity of American society. It is the only full-probability, personal-interview survey designed to monitor changes in both social characteristics and attitudes currently being conducted in the United States.

Links: <https://gss.norc.org/get-the-data>

A7 - Health Behaviors

Datasets that include data on health behaviors.

Behavioral Risk Factor Surveillance System (BRFSS): *State-level data on health-related risk behaviors, chronic health conditions, and use of preventive services*

Source: Centers for Disease Control, with data from the Behavioral Risk Factor Surveillance System (BRFSS)

Description:

The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

Links: <https://www.cdc.gov/brfss/>

National Survey on Drug Use and Health (NSDUH): *Health behaviors that contribute to the leading causes of death, disability, and social problems among youth and adults*

Source: Centers for Disease Control

Description:

The YRBSS was developed in 1990 to monitor health behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and adults in the United States. These behaviors, often established during childhood and early adolescence, include: behaviors that contribute to unintentional injuries and violence; sexual behaviors related to unintended pregnancy and sexually transmitted infections, including HIV infection; alcohol and other drug use; tobacco use; unhealthy dietary behaviors; inadequate physical activity. In addition, the YRBSS monitors the prevalence of obesity and asthma and other health-related behaviors plus sexual identity and sex of sexual contacts.

Links: <https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2019-nsduh-2019-ds0001>

Youth Risk Behavior Surveillance System (YRBSS): *Use of illicit drugs, alcohol, and tobacco and mental health issues*

Source: Centers for Disease Control

Description:

The NSDUH series, formerly the National Household Survey on Drug Abuse, is the leading source of statistical information on the use of illicit drugs, alcohol, and tobacco and mental health issues in the United States. The survey tracks trends in specific substance use and mental illness measures and assesses substance use disorders and treatment for these disorders. The population of the NSDUH series is the general civilian population aged 12 and older. Questions include age at first use, as well as lifetime, annual, and past-month use of the following drugs: alcohol, marijuana, cocaine (including crack), hallucinogens, heroin, inhalants, tobacco, pain relievers, tranquilizers, stimulants, and sedatives. The survey covers substance use treatment history and perceived need for treatment, and it includes questions from the "Diagnostic and Statistical Manual (DSM) of Mental Disorders" (DSM) that allow diagnostic criteria to be applied. Respondents are also asked about personal and family income, health care access and coverage, illegal activities and arrest records, problems resulting from the use of drugs, and perceptions of risks. Demographic data include gender, race,

age, ethnicity, educational level, employment status, income level, veteran status, household composition, and population density.

Links: <https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2019-nsduh-2019-ds0001>

B - HEALTH OUTCOMES

B1 - Diseases and Conditions

Datasets that include data on incidence and prevalence of specific diseases and health conditions.

CDC Births Data Summary: *Nativity data from CDC births*

Source: Centers for Disease Control

Description:

This dataset contains natality data based on CDC-collected statistics for live births occurring within the United States to U.S. residents.

The data capture a range of maternal demographic information, such as state and county of residence, mother's age and race, ethnicity and country of origin, marital status, and education. It also includes health and medical data on these mothers, including prior birth history, prenatal care visits, WIC enrollment, tobacco use, method of delivery, method of payment, and congenital anomalies and other morbidity data.

Beyond maternal characteristics, this dataset also illustrates both paternal and infant information that may be relevant to understanding certain social determinants of health. Paternal characteristics include age, race and ethnicity (including country of origin), education. Infant characteristics: gender, birth weight, delivery, congenital abnormalities.

For researchers and population health teams, this data can be used to identify localities that have had higher-than-average complicated and high-cost births and give insight into possible targeting strategies based on population characteristics. The data are derived from birth certificates, reported to the CDC.

Links: <https://console.cloud.google.com/marketplace/product/center-disease-control/wonder-births>

COVID-19 Data Repository by CSSE at JHU: *Repository of aggregated coronavirus COVID-19 cases by JHU*

Source: Johns Hopkins University

Description:

This is the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). The data include the location and number of confirmed COVID-19 cases, deaths, and recoveries for all affected countries, aggregated at the appropriate province/state. It was developed to enable researchers, public health authorities and the general public to track the outbreak.

For publications that use the data, please cite the following publication: Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis.* 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1

Links: https://console.cloud.google.com/marketplace/product/johnshopkins/covid19_jhu_global_case

COVID-19 Mobility Impact: *Changes in community movement due to COVID-19*

Source: Geotab

Description:

This dataset aims to provide insights into what has changed in response to policies aimed at combating COVID-19. It reports movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.

Links: <https://console.cloud.google.com/marketplace/product/geotab-public-data/covid19-mobility-impacts>

COVID-19 Open Data: *Daily time-series data related to COVID-19 globally*

Source: Google BigQuery Public Datasets Program

Description:

This repository contains the largest COVID-19 epidemiological database available in addition to a powerful set of expansive covariates. It includes open sourced data with a permissive license (enabling commercial use) relating to vaccinations, epidemiology, hospitalizations, demographics, economy, geography, health, mobility, government response, weather, and more. Moreover, the data merges daily time-series from hundreds of data sources at a fine spatial resolution, containing over 20,000 locations and using a consistent set of region keys.

This dataset is available in both the US and EU regions of BigQuery: COVID-19 Open Data: US Region COVID-19 Open Data: EU Region

All data in this dataset is retrieved automatically. When possible, data is retrieved directly from the relevant authorities, like a country's ministry of health.

Links: <https://console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-open-data>

COVID-19 Vaccination Access: *Metrics quantifying travel times to COVID-19 vaccination sites*

Source: Google BigQuery Public Datasets Program

Description:

The COVID-19 Vaccination Access Dataset characterizes access to COVID-19 vaccination sites based on travel times.

This dataset is intended to help public health officials, researchers, and healthcare providers to identify areas with insufficient access, deploy interventions, and research these issues—you shouldn't use this dataset for other purposes.

To learn more about the dataset and how we generate it, read the dataset documentation. To visualize the data, try exploring the Vaccine Equity Planner tool by Ariadne Labs and Boston Children's Hospital , which is powered by this dataset.

Links: <https://console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-vaccination-access>

U.S. Chronic Disease Indicators (CDI): *124 chronic disease indicators important to public health practice*

Source: Centers for Disease Control

Description:

CDC's Division of Population Health provides cross-cutting set of 124 indicators that were developed by consensus and that allows states and territories and large metropolitan areas to uniformly define, collect, and report chronic disease data that are important to public health practice and available for states, territories and large metropolitan areas. Examples include mortality due to diabetes and asthma hospitalizations. In addition to providing access to state-specific indicator data, the CDI web site serves as a gateway to additional information and data resources.

Links: <https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-CDI-/g4ie-h725>

National Mental Health Services Survey (N-MHSS): *Mental health treatment facilities and patients*

Source: Substance Abuse and Mental Health Services Administration (SAMHSA)

Description:

The National Mental Health Services Survey (N-MHSS) is designed to collect information from all specialty mental health facilities in the United States, both public and private. N-MHSS provides the mechanism for quantifying the dynamic character and composition of the U.S. mental health treatment delivery system. It is an annual survey designed to collect statistical information on the numbers and characteristics of all known mental health treatment facilities within the 50 states, the District of Columbia and the U.S. territories. In every other (even) year, beginning in 2014, the survey also collects statistical information on the numbers and demographic characteristics of persons served in these treatment facilities as of a specified survey reference date.

Links: <https://www.samhsa.gov/data/data-we-collect/nmhss-national-mental-health-services-survey>

National Survey of Substance Abuse Treatment Services (N-SSATS): *Substance abuse treatment facilities*

Source: Substance Abuse and Mental Health Services Administration (SAMHSA)

Description:

The National Survey of Substance Abuse Treatment Services (N-SSATS) is an annual census designed to collect information from all facilities within the 50 States, the District of Columbia, and the U.S. territories, both public and private, that provide substance abuse treatment. N-SSATS provides the mechanism for

quantifying the dynamic character and composition of the United States substance abuse treatment delivery system.

Links: <https://www.samhsa.gov/data/data-we-collect/n-ssats-national-survey-substance-abuse-treatment-services>

Disability and Health Data System (DHDS): *State-level data from BRFSS on adults with six functional disability types*

Source: Centers for Disease Control, with data from the Behavioral Risk Factor Surveillance System (BRFSS)

Description:

This online source of state-level data on adults with disabilities helps users access information on six functional disability types: cognitive, hearing, mobility, vision, self-care, and independent living. Data on more than 30 health topics among adults, with or without disabilities, can be explored in DHDS, including smoking, obesity, heart disease, and diabetes. Users can customize data maps, charts, and tables, making it easy to see information about their state or region. They can identify health differences between adults with and without disabilities overall, and by age, gender, and race/ethnicity.

Links: <https://data.cdc.gov/Disability-Health/Disability-and-Health-Data-System-DHDS-/k62p-6esq>

National Immunization Survey-Child (NIS-Child): *Coverage with all recommended vaccines among children aged 19-35 months*

Source: Centers for Disease Control

Description:

The National Immunization Survey-Child (NIS-Child) is a continuing nationwide telephone sample survey to monitor coverage with all recommended vaccines among children aged 19-35 months living in the United States and selected U.S. territories. The NIS-Child has been conducted annually since 1994.

Links: <https://www.cdc.gov/vaccines/imz-managers/nis/data-tables.html>

National Immunization Survey-Teen (NIS-Teen): *Coverage with all recommended vaccines among adolescents aged 13-17 years*

Source: Centers for Disease Control

Description:

The National Immunization Survey-Teen (NIS-Teen) is a continuing nationwide telephone sample survey to monitor coverage with all recommended vaccines among adolescents aged 13-17 years living in the United States and selected U.S. territories. The NIS-Teen has been conducted annually since 2006.

Links: <https://www.cdc.gov/vaccines/imz-managers/nis/data-tables.html>

B2 - Clinical Data/EHRs

Datasets that include data from electronic health records.

- No datasets are currently available in this category.