Congressional District Health Dashboard Downloadable Data Codebook

Updated February 20, 2024

Table of Contents

SECTION 1: Overview

The Congressional District Health Dashboard (the Dashboard) is a resource allowing users to view and compare data from multiple sources on health and the factors that shape health to guide policy solutions. The Dashboard selected over 30 metrics spanning five domains – clinical care, health behavior, health outcomes, physical environment, and social and economic factors – to quantify health, health determinants, and equity at the congressional district level.

Metrics are derived from both private and publicly available data sources, with some data sources contributing several metrics and others contributing only a single metric.

Document Mission

This document is written for an audience interested in using the data available for download from the Dashboard. It should be used in conjunction with the Technical Document, which outlines technical attributes of the data sources used on the Dashboard. The technical document can be found at https://congressionaldistricthealthdashboard.org/technical-documentation.

Users are invited to contact the Dashboard (info@cdhealthdashboard.org) with general feedback or questions not addressed below.

Downloading Dashboard Data

Dashboard data are available for free download at https://congressionaldistricthealthdashboard.org/data-access. Data are available for download in CSV and .txt formats.

Citing Dashboard Data

The Dashboard should be cited when the data or graphics are used, including in published presentations, articles, research, blogs, policy documents, and other print or digital media.

We encourage use of Dashboard data and visualizations, and suggest the following citation:

> Department of Population Health, NYU Langone Health. Congressional District Health Dashboard. https://congressionaldistricthealthdashboard.org/. Accessed [INSERT DATE OF ACCESS].

Data Updates

Dashboard data – both as displayed online and available for download – are updated periodically as underlying datasets are refreshed. Data release updates are announced by email and data versioning is tracked by the release date which is listed in the file name of the downloadable data files. Please email info@cdhealthdashboard.org with any questions or concerns.

Only data that are displayed on the Dashboard are available for download; older versions of Dashboard data are not available. Users in need of archived data should contact info@cdhealthdashboard.org.

Feedback or Errors

Users are encouraged to contact the Dashboard with comments or questions regarding the website and any documents available for download from the website, including this Codebook, at info@cdhealthdashboard.org.

SECTION 2: Key Notes on Downloadable Data

Data Available for Download

The Dashboard's data are accessible via downloadable CSV and .txt files. A list of metrics and years of data available for download are provided in Appendix A. The following data tables are available from the Dashboard:

Geographic Identifiers Lists

- List of all congressional districts presented on the Dashboard
- List of all tracts presented on the Dashboard

Datasets of All Metrics

- Metric data for all congressional districts presented on the Dashboard
- Metric data for all states presented on the Dashboard

Datasets of All Metrics, by State

- Metric data for all congressional districts presented on the Dashboard, separated by US state
- Metric data for all tracts presented on the Dashboard, separated by US state

Data Unavailable for Download

For Neighborhood Racial/Ethnic Segregation and Lead Exposure Risk Index, state values on the website are averages of congressional district values, rather than state values directly calculated or obtained from the data source. In these instances, we do not provide the state values for download, as these values can be calculated by averaging congressional district values. Please see the technical documentation for more information.

A full-country dataset for tract data is not provided due to file size constraints.

Multi-Year Data

All years of data available on the Dashboard are included in the same file for download. The year(s) of data available for each metric are outlined in Appendix A.

CSV and .txt File Structure

Data are provided in both CSV and .txt file format within the downloadable zipped file. All corresponding CSV and .txt files have identical contents. The Dashboard provides both CSV and .txt file formats for download to maximize flexibility for users.

Users are cautioned that leading zeroes are dropped in CSV files but not .txt files. For example, a FIPS code of 01 will appear as 1 in a CSV file and 01 in its corresponding .txt file; a FIPS code of 040703 will appear as 40703 in a CSV file and 040703 in its corresponding .txt file.

Metrics with Demographic Breakdown

Metric data for specific demographic groups available on the website (in the section "Metrics by Demographics") are also released in downloadable data.

Technical Documentation

Users are strongly encouraged to review the Technical Documentation available at
https://congressionaldistricthealthdashboard.org/technical-documentation in conjunction with this
Codebook prior to data analysis.

Note on Geographic Identifiers

Congressional districts are uniquely identified by concatenating state FIPS codes (Federal Information
Processing Series) and congressional district IDs to create a 4-digit unique identifier "[state FIPS code]
[congressional district ID]". For example, the geographic identifier for Alabama 2nd district is 0102.

Some states have only a single congressional district because their populations do not meet the
necessary size threshold for multiple districts. These states are considered "at-large" districts. For
internal consistency, the Dashboard uses "[state FIPS code]01" to designate these districts, as opposed
to the Census Bureau's designation of "[state FIPS code]00".

Tracts are identified by concatenating state, county, and tract FIPS codes to create an 11-digit identifier
"[state FIPS code][county FIPS code][tract FIPS code]". In the tract datasets, the variable parent_level
represents the unique identifier of the congressional district in which that tract resides. Some tracts
cross the boundaries of more than one congressional district. In this case, Dashboard data reports values
for each tract-congressional district combination as separate rows, causing duplication of the geo_fips
variable within Dashboard tract datasets. To establish unique identifiers within downloadable data, use
the geo_fips and parent_fips variables in conjunction.

Note on At-Large Districts

There are seven states that only contain a single congressional district (known as at-large districts). For
these states the congressional district and state values are identical.

Note on Versioning

The following entities are versioned by the date of the most recent data release when that entity was
updated. Versioning of each entity is independent, and not all will necessarily be updated each data
release.

1. Data
2. Technical Document
3. Codebook

The following entity is versioned by the Congress currently represented on the Dashboard (i.e. 118 for
the 118th Congress).

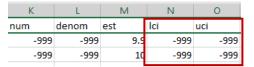1. Geographic Identifiers List

SECTION 3: Coding of Variables

Not Applicable versus Unavailable/Censored

Our data variables distinguish between (a) data that are not released because the value was *not applicable* to the metric and (b) data that are missing because they were *censored or unavailable in an underlying dataset*.

Not Applicable is Represented by -999

An entry of -999 indicates "not applicable".

For example, a value of -999 under the variables "lci" and "uci" means that confidence intervals are not calculated for this value (see Example A).

num	denom	est	lci	uci
-999	-999	9.9	-999	-999
-999	-999	10	-999	-999

Example A: Screenshot of CSV output

Here, only the estimate is relevant for the metric. The numerator, denominator, and confidence intervals are not applicable, as indicated by the -999 values.

Unavailable/Censored is Represented by NA

A value of *NA* indicates that there is no associated value for that variable. For the "est" variable a value of *NA* means the estimate was unavailable in an underlying dataset OR was censored by the Dashboard. The variable "data_indicator" gives additional information about unavailable or censored data (see section below). For other variables (i.e. "data_indicator") a value of *NA* means there is no additional information relevant for that estimate.

For example, a value of "*NA*" under the variable "est" and a value of "censor_2" under the variable "data_indicator" means that this estimate is censored (see Example B).

group_name	est	data_period	period_type	census_parent_shape_year	state_indicator	data_indicator
Other	NA	2020	3 Year Estimate	118	NA	censor_2
Total	22.1	2020	1 Year Estimate	118	NA	NA

Example B: Screenshot of CSV output

Here, the estimate is not released due to censorship.

Data Indicators

Additional information about state and congressional district data can be found under the variables "data_indicator", "state_indicator", and "agg_indicator". Data indicators are not provided for tract data.

The following values may be present under "data_indicator":

- censor_1 = Estimate was suppressed by the Dashboard team due to missing contributing data.
- censor_2 = Estimate was suppressed by the Dashboard team due to an insufficient number of contributing events, in accordance with National Center for Health Statistics suppression criteria.
- censor_3 = Estimate was suppressed by the Dashboard team due to insufficient sample size.
- censor_4 = Estimate was suppressed by the Dashboard team due to concerns about underlying data reliability for the state.
- cautions_1 = Use caution when interpreting this estimate, as some contributing data are missing which may impact its accuracy.
- cautions_2 = Use caution when interpreting this estimate, as underlying events are rare which may impact the estimate.
- missing_1 = Data are unavailable for this congressional district or state.

The following values may be present under "state_indicator":

- state_1 = Because this state only has one congressional district, the Dashboard presents the state-level value for this congressional district.

The following values may be present under "agg_indicator":

- aggregation_1 = This estimate is calculated by aggregating estimates from smaller geographies.

Please consult the technical document for more detail on data censorship and data source information.

Data Time Periods

The Dashboard includes metrics with different types of data time periods including data for a single year, pooled data from multiple years, monthly data, and quarterly data. The variable "data_period" represents the year, month, or quarter of the data used in calculation with the following formats:

- Year: formatted as yyyy, i.e. 2022
- Month: formatted as yyyymm, i.e. 202201
- Quarter: formatted as yyyymm with mm representing the first month of the quarter, i.e., 202210 represents the fourth quarter of 2022 spanning October-December 2022

Additional information about the data period type can be found under the variable "period_type".

The following values may be present under "period_type":

- 1 year estimate = Estimates are calculated based on data from a single year
- 3 year estimate = Estimates are calculated based on data from a 3-year period
- 5 year estimate = Estimates are calculated based on data from a 5-year period
- 6 year estimate = Estimates are calculated based on data from a 6-year period
- 1 year modeled estimate = Estimates are modeled based on data from a single year
- 6 year modeled estimate = Estimates are modeled based on data from a 6-year period
- School year ending in = Estimates are from the school year ending in the specified year (i.e. 2022 means from the 2021-2022 school year)
- 1 month average estimate = Estimates are averaged based on data from a 1-month period
- 1 quarter average estimate = Estimates are averaged based on data from a 3-month period
- -999 = Estimates are from the year specified in "data_period" with no additional relevant details

Please consult the technical document for more detail on these calculations.

Data Sources

For some metrics, data sources differ by geography-level. The variable "source_name" provides the source name abbreviation for metrics in each dataset. Full source names can be found in Appendix B.

Geographic Boundaries

For a given metric and year, the "census_parent_shape_year" variable specifies which census boundaries (2010 or 2020) are used for tracts (where applicable), and which congress boundaries are used for congressional districts. This variable is not applicable for states as the boundaries do not change across census years or congress sessions.

Note that to maintain consistency across the website, minor Census tract changes that occurred between 2010 and 2019 (and were differentially incorporated by data sources) are standardized back into 2010 Census tract geographies on the Dashboard.[1]

The following values may be present under "census_parent_shape_year":

- 118 = 118th Congress
- 2010_118 = tracts use 2010 census boundaries; CDs use the 118th Congress boundaries
- 2020_118 = tracts use 2020 census boundaries; CDs use the 118th Congress boundaries

SECTION 4: File Layout for Geographic Identifiers List

Geographic identifiers lists are available as CSV and .txt files with the names:

- CD_geographic_identifiers_[census_parent_shape_year]
- Tract_CD_geographic_identifiers_[census_parent_shape_year]

Variable Information

Variable Name	Variable Definition	Variable Type	Notes
geo_fips	FIPS code for the congressional district or census tract	*numeric*	4 digits for CD 11 digits for tract *(note that CSV drops leading zero)*
geo_name	Name of the congressional district or tract • ex: Ohio 2nd, 054689	*string*	Tracts are named with 6-digit codes *(note that CSV drops leading zero)*
geo_level	The geographic level of geo_fips • ex: cd, tract	*string*	cd = congressional district
parent_fips	FIPS code for the congressional district in which the tract resides	*numeric*	*Only included in tract file*
parent_name	Name of the congressional district in which the tract resides	*string*	*Only included in tract file*
parent_level	The geographic level of parent_fips • ex: cd	*string*	*Only included in tract file*
state_fips	State FIPS code	*numeric*	2 digits *(note that CSV drops leading zero)*
state_abbr	State abbreviation	*string*	
state_name	State name	*string*	
census_parent_shape_year	Year of the census and/or the Congress that correlates with geographic boundaries • ex: 118, 2010_118	*string*	See *Coding of Variables: Geographic Boundaries* section for details

SECTION 5: File Layout for Congressional District Data

Congressional district metric data are available as CSV and .txt files with the names
"All_states_CD_[version]" and "[state_abbr]_CD_[version]" when files are downloaded by state *(ex. "AL_CD_02-20-2024" for Alabama)*.

Variable Information

Variable Name	Variable Definition	Variable Type	Notes
state_abbr	State abbreviation	*string*	
state_fips	State FIPS code	*numeric*	2 digits *(note that CSV drops leading zero)*
geo_fips	FIPS code for the congressional district	*numeric*	4 digits *(note that CSV drops leading zero)*
geo_level	The geographic level of geo_fips • ex: cd	*string*	cd = congressional district
geo_name	Name of the congressional district • ex: Ohio 2nd	*string*	
metric_name	Metric name	*string*	
group_name	Demographic group	*string*	
est	Estimate	*numeric*	*NA* = censored or unavailable value
data_period	Time period of data	*numeric*	See *Coding of Variables: Data Time Periods* section for details
period_type	Estimate type • ex: 5 year estimate	*string*	See *Coding of Variables: Data Time Periods* section for details
source_name	Source name abbreviation	*string*	See *Coding of Variables: Data Sources* section for details
census_parent_shape_year	Year of either the Congress that correlates with geographic boundaries • ex: 118	*numeric*	See *Coding of Variables: Geographic Boundaries* section for details
state_indicator	Information about state data source	*string*	See *Coding of Variables: Data Indicators* section for details
data_indicator	Information about unavailable data	*string*	See *Coding of Variables: Data Indicators* section for details
agg_indicator	Information about data aggregation	*string*	See *Coding of Variables: Data Indicators* section for details

NOTE: num, denom, lci, and uci variables are not provided in congressional district-level datasets because these variables are not available for congressional districts at this time.

SECTION 6: File Layout for Tract Data

Tract metric data are available as CSV and .txt files with the name "[state_abbr]_Tract_CD_[version]"
when files are downloaded by state *(ex. "AL_Tract_CD_02-20-2024" for Alabama).*

Variable Information

Variable Name	Variable Definition	Variable Type	Notes
state_abbr	State abbreviation	*string*	
state_fips	State FIPS code	*numeric*	2 digits *(note that CSV drops leading zero)*
geo_fips	FIPS code for the tract	*numeric*	11 digits *(note that CSV drops leading zero)*
geo_level	The geographic level of geo_fips • ex: tract	*string*	
geo_name	Name of the tract • ex: 020100	*numeric*	Tracts are named with 6-digit codes *(note that CSV drops leading zeros)*
parent_fips	FIPS code for the congressional district in which the tract resides	*numeric*	4 digits *(note that CSV drops leading zero)*
parent_level	The geographic level of parent_fips • ex: cd	*string*	cd = congressional district
parent_name	Name of the congressional district in which the tract resides • ex: Ohio 2nd	*string*	
metric_name	Metric name	*string*	
group_name	Demographic group	*string*	
num	Numerator	*numeric*	-999 = not applicable NA = censored or unavailable value
denom	Denominator	*numeric*	-999 = not applicable NA = censored or unavailable value
est	Estimate	*numeric*	*NA* = censored or unavailable value
lci	Lower confidence interval	*numeric*	-999 = not applicable NA = censored or unavailable value
uci	Upper confidence interval	*numeric*	-999 = not applicable NA = censored or unavailable value
data_period	Time period of data	*numeric*	See *Coding of Variables: Data Time Periods* section for details
period_type	Estimate type • ex: 5 year estimate	*string*	See *Coding of Variables: Data Time Periods* section for details
source_name	Source name abbreviation	*string*	See *Coding of Variables: Data Sources* section for details
census_parent_shape_year	Year of the census and the Congress that correlates with geographic boundaries • ex: 2010_118	*string*	See *Coding of Variables: Geographic Boundaries* section for details

SECTION 7: File Layout for State Data

State metric data are available as CSV and .txt files with the names "All_states_State_[version]".

Variable Information

Variable Name	Variable Definition	Variable Type	Notes
state_abbr	State abbreviation	*string*	
geo_fips	FIPS code for the state	*numeric*	2 digits *(note that CSV drops leading zero)*
geo_level	The geographic level of geo_fips • ex: state	*string*	
geo_name	Name of the state	*string*	
metric_name	Metric name	*string*	
group_name	Demographic group	*string*	
num	Numerator	*numeric*	-999 = not applicable NA = censored or unavailable value
denom	Denominator	*numeric*	-999 = not applicable NA = censored or unavailable value
est	Estimate	*numeric*	*NA* = censored or unavailable value
lci	Lower confidence interval	*numeric*	-999 = not applicable NA = censored or unavailable value
uci	Upper confidence interval	*numeric*	-999 = not applicable NA = censored or unavailable value
data_period	Time period of data	*numeric*	See *Coding of Variables: Data Time Periods* section for details
period_type	Estimate type • ex: 5 year estimate	*string*	See *Coding of Variables: Data Time Periods* section for details
source_name	Source name abbreviation	*string*	See *Coding of Variables: Data Sources* section for details
data_indicator	Information about unavailable data	*string*	See *Coding of Variables: Data Indicators* section for details
agg_indicator	Information about data aggregation	*string*	See *Coding of Variables: Data Indicators* section for details

NOTE: state_indicator variable is not provided in state-level datasets because this indicator is not relevant to state data.

SECTION 8: Data Update Log

Version	Update Notes
02-20-2024	• Medicaid Enrollment - % of Total Population: New metric added (Q4 2022 – Q3 2023) • ACS: o Addition of 2022 data o Calculation update for Lead Exposure Risk Index and Housing with Potential Lead Risk • PLACES Project: o Addition of 2021 data for all metrics except Dental Care and Preventative Services, 65+ o Calculation update for 2020 state Preventive Services, 65+ data • NVSS: Addition of 2021 data • Chronic Absenteeism: New metric added (School year ending in 2022) • Air Pollution – Ozone: New monthly metric added (Jan – Dec 2022) • Air Pollution – PM2.5: o Update of data source o Change in data time period from annual to monthly data (Jan – Dec 2022) • COVID Local Risk Index: Metric removed from website and downloadable data
06-06-2023	• 08-23-2023 revision: corrected an error in segregation 2021 congressional district and state values that caused a slight inflation of scores. • ACS: Addition of 2021 data • Air Pollution – Particulate Matter: Addition of 2019 data • PLACES Project: o Addition of 2020 data for all metrics except High Blood Pressure o Addition of 2017 data for High Blood Pressure • Removal of Limited Supermarket Proximity • Addition of site-wide tract data, where available • Addition of site-wide state data, where available
03-10-2023	• Limited Supermarket Proximity: Metric name updated from "Limited Access to Healthy Foods"
01-24-2023	First release of the Congressional District Health Dashboard • 36 metrics • Data for all congressional districts • One year of data released for all metrics

SECTION 9: Notes on Specific Metrics

Life expectancy

The Dashboard reports tract-level data as received from National Center for Health Statistics/USALEEP; tract-level data and documentation files are available for free download.[2] A flag to indicate the source of the age-specific death rates used to calculate the abridged period life tables is included in the data downloaded directly from USALEEP. However, this variable is ***not*** included in the Dashboard downloadable data. Users requiring this information should refer to the National Center for Health Statistics/USALEEP dataset.

NVSS metrics

Please note that the data period is different between demographic subgroups (3-year estimates) and total population (1-year estimates). Users may notice different trends across subgroups and total population over time due to the difference in data period. Please keep this difference in mind when making comparisons across demographic subgroups and total population estimates.

SECTION 10: References

1.   US Census Bureau. Table & Geography Changes. https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes.html

2.   National Center for Health Statistics. U.S. Small-Area Life Expectancy Estimates Project - USALEEP: Life Expectancy Estimates File for United States, 2010-2015. National Center for Health Statistics,. Updated June 9, 2020. Accessed February 6, 2024. https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html

APPENDIX A:

Data Period Availability, by Metric

Metric Name	Data Periods Available
Air Pollution - Ozone	01/2022-12/2022*
Air Pollution - Particulate Matter	01/2022-12/2022*
Binge Drinking	2019, 2020, 2021
Breast Cancer Deaths	2020, 2021
Broadband Connection	2020, 2021, 2022
Cardiovascular Disease Deaths	2020, 2021
Children in Poverty	2020, 2021, 2022
Chronic Absenteeism	2022
Colorectal Cancer Deaths	2020, 2021
Dental Care	2018, 2020
Diabetes	2019, 2020, 2021
Firearm Homicides	2020, 2021
Firearm Suicides	2020, 2021
Frequent Mental Distress	2019, 2020, 2021
Frequent Physical Distress	2019, 2020, 2021
High Blood Pressure	2017, 2019, 2021
High School Completion	2020, 2021, 2022
Housing with Potential Lead Risk	2020, 2021, 2022
Income Inequality	2020, 2021, 2022
Lead Exposure Risk Index	2020, 2021, 2022
Life Expectancy	2015
Low Birthweight	2020, 2021
Medicaid Enrollment - % of Total Population	10/2022-07/2023**
Neighborhood Racial/Ethnic Segregation	2020, 2021, 2022
Obesity	2019, 2020, 2021
Opioid Overdose Deaths	2020, 2021
Physical Inactivity	2019, 2020, 2021
Premature Deaths (All Causes)	2020, 2021
Prenatal Care	2020, 2021
Preventive Services, 65+	2018, 2020
Racial/Ethnic Diversity	2020, 2021, 2022
Rent Burden	2020, 2021, 2022
Routine Checkup, 18+	2019, 2020, 2021
Smoking	2019, 2020, 2021
Teen Births	2020, 2021
Unemployment	2020, 2021, 2022
Uninsured	2020, 2021, 2022

* Monthly data   **Quarterly data

See *Coding of Variables: Data Time Periods* section for more details on data period formatting.

APPENDIX B:

Data Sources

Source Name Abbreviation	Full Source Name
ACS	American Community Survey, U.S. Census Bureau
NVSS ND	Natality Data, National Vital Statistics System, National Center for Health Statistics
NVSS MCDD	Multiple Cause of Death Data, National Vital Statistics System, National Center for Health Statistics
USALEEP	U.S. Small-Area Life Expectancy Estimates Project (USALEEP)
NCES	National Center for Education Statistics, U.S. Department of Education
PLACES	PLACES Project, Centers for Disease Control
GMU	George Mason University
BRFSS	Behavioral Risk Factor Surveillance System
ACS CMS	American Community Survey and Centers for Medicare & Medicaid Services