Congressional District Health Dashboard Downloadable Data Codebook

Updated January 24, 2023

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 behaviors, 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 and in the "version" variable in 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 the Appendix. The following data tables are available from the Dashboard:

Geographic Identifiers List

- List of all congressional districts represented on the Dashboard

Dataset of All Metrics

- Metric data for all congressional districts represented on the Dashboard

Datasets of All Metrics, by State

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

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".

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).

	K	L	M	N	O
	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" or "period_type") 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).

G	H	I	J	K	L	M
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 the data can be found under the variables "data_indicator" and "state_indicator."

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 NCHS suppression criteria.
- 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.

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.

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

Data Time Periods

Some metrics are pooled estimates combining multiple years of data. "data_period" represent the most recent year of the data used in calculation. 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
- 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
- -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.

SECTION 4: File Layout for Geographic Identifiers List

A geographic identifiers list is available as CSV and .txt files with the name "CD_geographic_identifiers".

Variable Information

Variable Name	Variable Definition	Variable Type	Notes
geo_fips	FIPS code for the congressional district	*numeric*	4 digits (note that CSV drops leading zero)
geo_name	Name of the congressional district • ex: Ohio 2nd	*string*	
geo_level	The geographic level of geo_fips • ex: cd	*string*	cd = congressional district
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 either the census or the Congress that correlates with geographic boundaries • ex: 118	*numeric*	

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_01-24-2023" 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	Year 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
census_parent_shape_year	Year of either the census or the Congress that correlates with geographic boundaries • ex: 118	*numeric*	
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
version	Date of data update	*string*	

SECTION 6: Data Update Log

Version	Update Notes
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

APPENDIX:

Data Year Availability, by Metric

Metric Name	Years Available	Estimate Type
Air Pollution - Particulate Matter	2018	Not Applicable
Binge Drinking	2019	1 Year Modeled Estimate
Breast Cancer Deaths	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Broadband Connection	2020	5 Year Estimate
Cardiovascular Disease Deaths	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Children in Poverty	2020	5 Year Estimate
Colorectal Cancer Deaths	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Covid Local Risk Index	2022	Data from July, 2019 and 2015-2019 Estimates
Dental Care	2018	1 Year Modeled Estimate
Diabetes	2019	1 Year Modeled Estimate
Firearm Homicides	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Firearm Suicides	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Frequent Mental Distress	2019	1 Year Modeled Estimate
Frequent Physical Distress	2019	1 Year Modeled Estimate
High Blood Pressure	2019	1 Year Modeled Estimate
High School Completion	2020	5 Year Estimate
Housing With Potential Lead Risk	2020	5 Year Estimate
Income Inequality	2020	5 Year Estimate
Lead Exposure Risk Index	2020	5 Year Estimate
Life Expectancy	2015	6 Year Modeled Estimate
Limited Access to Healthy Foods	2019	Not Applicable
Low Birthweight	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Neighborhood Racial/Ethnic Segregation	2020	5 Year Estimate
Obesity	2019	1 Year Modeled Estimate
Opioid Overdose Deaths	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Physical Inactivity	2019	1 Year Modeled Estimate
Premature Deaths (All Causes)	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Prenatal Care	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Preventive Services, 65+	2018	1 Year Modeled Estimate

Metric Name	Years Available	Estimate Type
Racial/Ethnic Diversity	2020	5 Year Estimate
Rent Burden	2020	5 Year Estimate
Routine Checkup, 18+	2019	1 Year Modeled Estimate
Smoking	2019	1 Year Modeled Estimate
Teen Births	2020	Total: 1 Year Estimate
		Subgroups: 3 Year Estimate
Unemployment	2020	5 Year Estimate
Uninsured	2020	5 Year Estimate