

[Logic Liaison Template][N3C Covid Enclave] Social Determinates of Health for All Patients README

ASSET CREATOR(s): Andrea Zhou (UVA) and Johanna Loomba (UVA) with design support from Charisse Madlock-Brown from the SDOH Domain Team and code review support from other members of the Logic Liaison team.

README SECTIONS:

TEMPLATE STATUS

PURPOSE & USE

DATA DICTIONARY

TEMPLATE LIMITATIONS

RELEASE NOTES (VERSION CHANGES)

TEMPLATE STATUS:

AS OF SEP, 2022: Initial version of the template produces a final person-level dataset that contains 4 county-level SDOH variables from the [BU-SC_SDOH_N3C_2018_20201130](#) external dataset and 66 ZCTA-level SDOH variables from the [ZCTA_by_SDOH_percentages](#) external dataset for all patients in the Enclave. See the release notes section in this document for additional details.

PURPOSE & USE:

This dataset provides N3C researchers with a curated set of 70 geographically-based SDOH measures tied to each patient. Because joining this data requires a valid five digit zip code, these fields are only available for patients with a five digit zip code in Level 3 (LDS) data. Any project approved for level 3 data can import this dataset, select the desired columns, and join to any table containing person_id in order to quickly add SDOH variables to their analysis.

The majority of the variables are from the external dataset [ZCTA_by_SDOH_percentages](#) and mapped using [ZiptoZcta_Crosswalk_2021_ziptozcta2020](#) as a crosswalk while the remaining variables were brought in from the external dataset [BU-SC_SDOH_N3C_2018_20201130](#) using the [1-1_ZIP_COUNTY_HUD_file](#) as a crosswalk. Patient_ids were joined

to SDOH data based on the ZCTA and Preferred_County_FIPS (majority residence), respectively, associated with the patient's zip code. The five digit zip, ZCTA, and Preferred_County_FIPS are listed for each patient in the dataset for easy re-use as needed.



Refer to the documentation linked below for further information:


[Documentation for BU-SC_SDOH_N3C_2018_20201130](#)


Documentation for ZCTA_by_SDOH_percentages ([Census Tables, 2019: ACS 5-Year Estimates Detailed Tables: P8, P2, S0101, B19001, S1701, B08006, B08135, B05001, B27001, B27002, B27003, S2301, B15002, B06007, B25014, B11011](#)) paired with [2019 ACS Subject Definitions](#)


DATA DICTIONARY:


Five_Digit_Zip, string: Five digit zip codes from [location](#) (subset of **postal_code** below)



location_id, string: Location identifier from  **person** used to link to the patient's information in  **location**


person_id, string: Unique identifier for the N3C patient from  **person**


data_partner_id, integer: The anonymized institution contributing this patient's record to N3C from  **person**


city, string: City from  **location**


state, string: State from  **location**


postal_code, string: Zip from  **location** **county**,
string: County from  **location**


ZCTA, string: ZCTA (ZIP Code Tabulation Areas) from  **ZiptoZcta_Crosswalk_2021_ziptozcta2020** used to map to **Five_Digit_Zip**


African_American, string: % of total population who identify as Black or African American alone from  **ZCTA_by_SDoH_percentages**

AIAN, string: % of total population who identify as American Indian and Alaskan Native alone from  **ZCTA_by_SDoH_percentages**


Asian, string: % of total population who identify as Asian alone from  **ZCTA_by_SDoH_percentages**


NHPI, string: % of total population who identify as Native Hawaiian and Other Pacific Islander alone from  **ZCTA_by_SDoH_percentages**


White, string: % of total population who identify as White alone from  **ZCTA_by_SDoH_percentages**

other_race, string: % of total population who identify as a race not listed in another single race category from  **ZCTA_by_SDoH_percentages**


two_or_more_races, string: % of total population who identify as more than one race from  **ZCTA_by_SDoH_percentages**


Hispanic_any_race, string: % of total population who identify as Hispanic or Latino, regardless of race from  **ZCTA_by_SDoH_percentages**. % of total population who identify as Not Hispanic or Latino would be the full complement to this value.


male, string: % of total population who identify as male from  **ZCTA_by_SDoH_percentages** **female**,


string: % of total population who identify as female from  **ZCTA_by_SDoH_percentages**


age_18_and_under, string: % of total population under the age of 18 from  **ZCTA_by_SDoH_percentages**


age_18_and_over, string: % of total population over the age of 18 from  **ZCTA_by_SDoH_percentages**


age_65plus, string: % of total population over the age of 65 from  **ZCTA_by_SDoH_percentages**


income_10k_and_under, string: % of households with income under \$10,000 annually from  **ZCTA_by_SDoH_percentages**


income_10k_15k, string: % of households with income between \$10,000 - \$14,999 annually from  **ZCTA_by_SDoH_percentages**


income_15k_20k, string: % of households with income between \$15,000 - \$19,999 annually from  **ZCTA_by_SDoH_percentages**

income_20k_25k, string: % of households with income between \$20,000 - \$24,999 annually from  **ZCTA_by_SDoH_percentages**

income_25k_30k, string: % of households with income between \$25,000 - \$29,999 annually from  **ZCTA_by_SDoH_percentages**

income_30k_35k, string: % of households with income between \$30,000 - \$34,999 annually from  **ZCTA_by_SDoH_percentages**

income_35k_40k, string: % of households with income between \$35,000 - \$39,999 annually from  **ZCTA_by_SDoH_percentages**

income_40k_45k, string: % of households with income between \$40,000 - \$44,999 annually from  **ZCTA_by_SDoH_percentages**

income_45k_50k, string: % of households with income between \$45,000 - \$49,999 annually from [ZCTA_by_SDoH_percentages](#)

income_50k_60k, string: % of households with income between \$50,000 - \$59,999 annually from [ZCTA_by_SDoH_percentages](#)

income_60k_75k, string: % of households with income between \$60,000 - \$74,999 annually from [ZCTA_by_SDoH_percentages](#)

income_75k_100k, string: % of households with income between \$75,000 - \$99,999 annually from [ZCTA_by_SDoH_percentages](#)

income_100k_125k, string: % of households with income between \$100,000 - 124,999 annually from [ZCTA_by_SDoH_percentages](#)

income_125k_150k, string: % of households with income between \$125,000 - \$149,999 annually from [ZCTA_by_SDoH_percentages](#)

income_150k_200k, string: % of households with income between \$150,000 - \$199,999 annually from [ZCTA_by_SDoH_percentages](#)

income_200k_and_over, string: % of households with income over \$200,000 annually from [ZCTA_by_SDoH_percentages](#)

poverty_status, string: % of population living at or below the poverty level where poverty status is determined based on poverty status of householder from [ZCTA_by_SDoH_percentages](#)

unemployment_status_age_16plus, string: % of population aged 16 and older who are unemployed from [ZCTA_by_SDoH_percentages](#)

unemployment_status_age_20to64, string: % of population aged 20 to 64 who are unemployed from [ZCTA_by_SDoH_percentages](#)

commute_drove, string: % of workers 16 years and older who drove alone or carpooled for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_drove_alone, string: % of workers 16 years and older who drove alone for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_drove_carpool, string: % of workers 16 years and older who carpooled for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_public_transit, string: % of workers 16 years and older who used public transit for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_bike, string: % of workers 16 years and older who rode a bike for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_walk, string: % of workers 16 years and older who walked for their transportation to work from [ZCTA_by_SDoH_percentages](#)

commute_work_from_home, string: % of workers 16 years and older who worked from home from [ZCTA_by_SDoH_percentages](#)

time_to_work_30min_plus, string: % of workers 16 years and older who did not work at home and had aggregate travel time to work of 30 minutes or more from [ZCTA_by_SDoH_percentages](#)

citizen_US, string: % of total population who are US citizens from [ZCTA_by_SDoH_percentages](#) **citizen_naturalized**, string: % of total population who are naturalized citizens from [ZCTA_by_SDoH_percentages](#) **citizen_non_US**, string: % of total population who are not US citizens from [ZCTA_by_SDoH_percentages](#)

speak_Spanish_and_speak_English_less_than_very_well, string: % of population 5 years and over who speak Spanish as their primary language and speak English less than "very well" from [ZCTA_by_SDoH_percentages](#)

speak_other_languages_and_speak_English_less_than_very_well, string: % of population 5 years and over who speak languages, other than English and Spanish, as their primary language and speak English less than "very well" from [ZCTA_by_SDoH_percentages](#)

owner_occupied_crowded_housing, string: % of owner-occupied housing units where occupants are living in crowded conditions. Crowded > 1 person per room from [ZCTA_by_SDoH_percentages](#)

renter_occupied_crowded_housing, string: % of renter-occupied housing units where occupants are living in crowded conditions. Crowded > 1 person per room from [ZCTA_by_SDoH_percentages](#)

owner_occupied_severely_crowded_housing, string: % of owner-occupied housing units where occupants are living in severely crowded conditions. Severely crowded > 1.5 persons per room from [ZCTA_by_SDoH_percentages](#)

renter_occupied_severely_crowded_housing, string: % of renter-occupied housing units where occupants are living in severely crowded conditions. Severely crowded > 1.5 persons per room from [ZCTA_by_SDoH_percentages](#)

family_married_couple, string: % of households with married couples with one or more individuals related to householder by birth, marriage, or adoption from [ZCTA_by_SDoH_percentages](#)

family_single_householder, string: % of household with a single householder with one or more individuals related to householder by birth, marriage, or adoption from [ZCTA_by_SDoH_percentages](#)

nonfamily_living_alone, string: % of households where a householder is living alone or with non-relatives only from [ZCTA_by_SDoH_percentages](#)

health_insurance, string: % of civilian noninstitutionalized population of any age with health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_19to64, string: % of civilian noninstitutionalized population aged 19-64 with health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_65plus, string: % of civilian noninstitutionalized population aged 65 and older with health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_private, string: % of civilian noninstitutionalized population of any age with private health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_private_19to64, string: % of civilian noninstitutionalized population aged 19-64 with private health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_private_65plus, string: % of civilian noninstitutionalized population aged 65 and older with private health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_public, string: % of civilian noninstitutionalized population of any age with public health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_public_19to64, string: % of civilian noninstitutionalized population aged 19-64 with public health insurance coverage from [ZCTA_by_SDoH_percentages](#)

health_insurance_public_65plus, string: % of civilian noninstitutionalized population aged 65 and older with public health insurance coverage from [ZCTA_by_SDoH_percentages](#)

education_hs_diploma, string: % of population 25 years and older with only a high school diploma (including equivalency) from [ZCTA_by_SDoH_percentages](#)

education_some_college, string: % of population 25 years and older with some college education from [ZCTA_by_SDoH_percentages](#)

education_associate_degree, string: % of population 25 years and older with an associate's degree from [ZCTA_by_SDoH_percentages](#)

education_bachelors_degree, string: % of population 25 years and older with a bachelor's degree from [ZCTA_by_SDoH_percentages](#)

education_graduate_or_professional_degree, string: % of population 25 years and older with a graduate or professional degree from [ZCTA_by_SDoH_percentages](#)

Preferred_County_FIPS, string: FIPS (Federal Information Processing Standard Publication) from [1-1_ZIP_COUNTY_HUD_file](#) used to represent majority residence of a zip code and map to **Five_Digit_Zip**

percent_voted_house_16_by_preferred_county, double: Percent of population who voted for a House representative in 2016 from [BU-SC_SDOH_N3C_2018_20201130](#)

social_community_score_by_preferred_county, double: SDoH Index v1.0 social and community overall score from [BU-SC_SDOH_N3C_2018_20201130](#)

MDs_by_preferred_county, double: Active non-federal MDs per 1,000 residents from [BU-SC_SDOH_N3C_2018_20201130](#)

sdoh2_by_preferred_county, double: SDoH Index v2.0 total score from [BU-SC_SDOH_N3C_2018_20201130](#)

TEMPLATE LIMITATIONS

This template's dataset only provides SDoH variables at ZCTA level granularity for 66 variables based on data from 2015-2019 and county level granularity for the remaining 4 variables based on data from 2018 that do not exist in the [ZCTA_by_SDoH_percentages](#) dataset produced by the SDoH domain team. The 4 county-level variables are joined in from [BU-SC_SDOH_N3C_2018_20201130](#).

The template's dataset also only provides this information for patients with a valid five digit zip code which is sometimes missing or obscured for privacy even in LDS data. "Preferred County" is derived from zip code by assigning the county that the majority of people in that zip code live in. This is true even if there is something listed in the county field because the string notation of counties is variable and thus not usable in our algorithms.

RELEASE NOTES:

TEMPLATE RELEASE NOTES (VERSION CHANGES): Sep

19, 2022:

Initial template publication