

## Codebook

Social Determinants of Health Impact on Quality of Life in US Counties Codebook  
By: Steven Uzupis, Udumaga Onyeukwu, and Miranda Gemme

<b>Dataset name:</b>	beta_codebook
<b>Dataset size:</b>	1.1 Mb
<b>Column count:</b>	47
<b>Row count:</b>	2,988
<b>Updated date:</b>	2024-07-30

### Description:

Team Beta capstone project for Merrimack College, Masters in Data Science Program, focusing on using Social Determinants of Health to predict quality of life in US Counties.

Research Question: How do social determinants of life affect the quality of life in different localities?

Hypothesis: Social determinants of health, such as economic stability, social connectedness, access to healthcare, and neighborhood environment, significantly predict self-reported health status in US counties.

Prediction: US counties with higher economic security, stronger social support infrastructure, better access to healthcare services, and safer, more accessible neighborhoods will report better overall health status than counties lacking these social determinants.

The predictor variables are obtained from the Social Determinants of Health (SDOH) Database provided by the Agency for Healthcare Research and Quality (AHRQ). The response variable represents 'the percentage of adults reporting fair or poor health.'

To test the research question or hypothesis, the statistical software R is being used.

### Column Attributes:

1	Column name:	fips_code		
	Column description:	State-county FIPS Code (5-digit)		
	Source information:	Both SDOH and CHR		
	Data type:	Numeric		
	Unique non-missing value count:	2,988		
	Missing value count:	0		
Min	Mean	Median	Max	SD
1001.00	30247.00	29118.00	56045.00	14882.03

<b>2</b>	<b>Column name:</b>	<b>weighted_population</b>
----------	---------------------	----------------------------

Column description: Total weighted population

Source information: SDOH

Data type: Numeric

Unique non-missing value count: 2,936

Missing value count: 0

Min	Mean	Median	Max	SD
645.00	103611.70	26410.50	10040682.00	331019.48

<b>3</b>	<b>Column name:</b>	<b>average_hh_size</b>
----------	---------------------	------------------------

Column description: Average household size

Source information: SDOH

Data type: Numeric

Unique non-missing value count: 161

Missing value count: 0

Min	Mean	Median	Max	SD
1.84	2.49	2.46	4.35	0.26

<b>4</b>	<b>Column name:</b>	<b>pct_male</b>
----------	---------------------	-----------------

Column description: Percentage of population that is male

Source information: SDOH

Data type: Numeric

Unique non-missing value count: 786

Missing value count: 0

Min	Mean	Median	Max	SD
41.99	50.07	49.62	69.54	2.35

<b>5</b>	<b>Column name:</b>	<b>pct_native_american</b>
----------	---------------------	----------------------------

Column description: Percentage of population reporting American Indian and Alaska Native race alone

Source information:	SDOH			
Data type:	Numeric			
Unique non-missing value count:	480			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.00	1.65	0.33	87.31	6.36

6	Column name:	pct_asian			
	Column description:	Percentage of population reporting Asian race alone			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	510			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	0.00	1.26	0.63	31.25	2.14

7	Column name:	pct_black			
	Column description:	Percentage of population reporting Black or African American ethnicity alone			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	1,369			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	0.00	9.01	2.34	87.79	14.32

8	Column name:	pct_hispanic
	Column description:	Percentage of population reporting Hispanic ethnicity
	Source information:	SDOH
	Data type:	Numeric

Unique non-missing value count:		1,449		
Missing value count:		0		
Min	Mean	Median	Max	SD
0.00	9.61	4.36	98.90	13.95

9	Column name:	pct_other_race			
	Column description:	Percentage of population reporting some other ethnicity alone			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	708			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	0.00	2.23	1.06	54.70	3.82

10	Column name:	pct_white			
	Column description:	Percentage of population reporting White race alone			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	1,968			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	9.35	82.36	88.26	99.90	16.00

11	Column name:	pct_single_parent
Column description:		Percentage of families with children that are single-parent families
Source information:		SDOH
Data type:		Numeric
Unique non-missing value count:		2,030

Missing value count:		0		
Min	Mean	Median	Max	SD
0.00	29.90	28.98	93.95	10.59

12	Column name:	pct_hh_other_computer			
	Column description:	Percentage of households with other type of computer			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	528			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	0.00	2.10	1.83	17.85	1.58

13	Column name:	pct_hh_internet			
	Column description:	Percentage of households with internet access			
	Source information:	SDOH			
	Data type:	Numeric			
	Unique non-missing value count:	1,833			
	Missing value count:	0			
	Min	Mean	Median	Max	SD
	41.38	78.95	80.10	96.81	8.05

14	Column name:	pct_employed		
	Column description:	Percentage of civilian labor force that is employed (ages 16 and over)		
	Source information:	SDOH		
	Data type:	Numeric		
	Unique non-missing value count:	895		
	Missing value count:	0		
Min	Mean	Median	Max	SD

69.61	94.81	95.12	100.00	2.47
-------	-------	-------	--------	------

<b>15</b>	<b>Column name:</b>	<b>pct_hh_inc_99999</b>
-----------	---------------------	-------------------------

Column description:	Percentage of population with household income between \$50,000 and \$99,999
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	1,373
Missing value count:	0

Min	Mean	Median	Max	SD
13.50	31.13	31.48	51.55	4.28

<b>16</b>	<b>Column name:</b>	<b>pct_w_medicare</b>
-----------	---------------------	-----------------------

Column description:	Percentage of population with Medicare
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	837
Missing value count:	0

Min	Mean	Median	Max	SD
1.75	6.51	6.19	19.44	2.16

<b>17</b>	<b>Column name:</b>	<b>clinical_nurse_pt</b>
-----------	---------------------	--------------------------

Column description:	Total number of clinical nurse specialists with NPI per 1,000 population
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	29
Missing value count:	0

Min	Mean	Median	Max	SD
0.00	0.01	0.00	0.41	0.03

<b>18</b>	<b>Column name:</b>		<b>dentist_pt</b>		
	Column description:		Total number of dentists with NPI per 1,000 population		
	Source information:		SDOH		
	Data type:		Numeric		
	Unique non-missing value count:		151		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	0.00	0.45	0.42	7.49	0.32
<b>19</b>	<b>Column name:</b>		<b>pa_pt</b>		
	Column description:		Total number of physician assistants with NPI per 1,000 population		
	Source information:		SDOH		
	Data type:		Numeric		
	Unique non-missing value count:		168		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	0.00	0.31	0.23	13.14	0.40
<b>20</b>	<b>Column name:</b>		<b>mental_health_faciliy_pt</b>		
	Column description:		Total number of facilities that provide mental health services per 1,000 population		
	Source information:		SDOH		
	Data type:		Numeric		
	Unique non-missing value count:		1,143		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	0.00	0.05	0.04	1.01	0.07
<b>21</b>	<b>Column name:</b>		<b>population_density</b>		

Column description:	Population density (County)
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	2,688
Missing value count:	0

Min	Mean	Median	Max	SD
0.44	211.10	44.97	71895.54	1472.62

<b>22</b>	<b>Column name:</b>	<b>days_over_90_f</b>
-----------	---------------------	-----------------------

Column description:	Number of days over 90 degrees Fahrenheit
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	144
Missing value count:	0

Min	Mean	Median	Max	SD
0.00	56.07	55.00	144.00	38.11

<b>23</b>	<b>Column name:</b>	<b>median_er_dist</b>
-----------	---------------------	-----------------------

Column description:	Median distance in miles to the nearest emergency department, calculated using population weighted tract centroids in the county
Source information:	SDOH
Data type:	Numeric
Unique non-missing value count:	1,297
Missing value count:	0

Min	Mean	Median	Max	SD
0.13	6.87	5.00	61.45	6.21

<b>24</b>	<b>Column name:</b>	<b>median_trauma_center_dist</b>
-----------	---------------------	----------------------------------



Column description:	Median distance in miles to the nearest designated trauma center, calculated using population weighted tract centroids in the county			
Source information:	SDOH			
Data type:	Numeric			
Unique non-missing value count:	2,172			
Missing value count:	0			

Min	Mean	Median	Max	SD
0.13	20.70	18.31	136.88	16.43

<b>25</b>	<b>Column name:</b>	<b>median_pediatric_icu_dist</b>		
-----------	---------------------	----------------------------------	--	--

Column description:	Median distance in miles to the nearest pediatric ICU, calculated using population weighted tract centroids in the county			
Source information:	SDOH			
Data type:	Numeric			
Unique non-missing value count:	2,501			
Missing value count:	0			

Min	Mean	Median	Max	SD
0.77	38.08	33.89	172.30	25.43

<b>26</b>	<b>Column name:</b>	<b>median_health_clinic_dist</b>		
-----------	---------------------	----------------------------------	--	--

Column description:	Median distance in miles to the nearest health clinic (FQHC, RHC), calculated using population weighted tract centroids in the county			
Source information:	SDOH			
Data type:	Numeric			
Unique non-missing value count:	1,033			
Missing value count:	0			

Min	Mean	Median	Max	SD
0.13	4.80	3.40	57.56	4.75

<b>27</b>	<b>Column name:</b>		<b>median_drug_alcohol_care_dist</b>		
	Column description:		Median distance in miles to the nearest hospital with alcohol and drug abuse inpatient care, calculated using population weighted tract centroids in the county		
	Source information:		SDOH		
	Data type:		Numeric		
	Unique non-missing value count:		2,221		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	0.61	23.41	20.02	148.80	19.21
<b>28</b>	<b>Column name:</b>		<b>percent_grandparents_as_guardians</b>		
	Column description:		Percentage of children living with grandparent householder whose grandparent is responsible for them: parent may or may not be present		
	Source information:		SDOH		
	Data type:		Numeric		
	Unique non-missing value count:		2,953		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	0.00	5.16	4.24	28.91	3.81
<b>29</b>	<b>Column name:</b>		<b>median_hh_income</b>		
	Column description:		Estimated median household income		
	Source information:		CHR		
	Data type:		Numeric		
	Unique non-missing value count:		2,895		
	Missing value count:		0		
	Min	Mean	Median	Max	SD
	28236.00	58599.57	56452.50	134050.00	14774.12

<b>30</b>	<b>Column name:</b>	<b>pct_poor_to_fair_health</b>
-----------	---------------------	--------------------------------

Column description: Percentage of adults reporting fair or poor health

Source information: CHR

Data type: Numeric

Unique non-missing value count: 226

Missing value count: 0

Min	Mean	Median	Max	SD
0.06	0.16	0.15	0.37	0.04

<b>31</b>	<b>Column name:</b>	<b>pct_adult_smokers</b>
-----------	---------------------	--------------------------

Column description: Percentage of adults who are current smokers

Source information: CHR

Data type: Numeric

Unique non-missing value count: 229

Missing value count: 0

Min	Mean	Median	Max	SD
0.07	0.20	0.20	0.41	0.04

<b>32</b>	<b>Column name:</b>	<b>pct_binge_drinkers</b>
-----------	---------------------	---------------------------

Column description: Percentage of adults reporting binge or heavy drinking

Source information: CHR

Data type: Numeric

Unique non-missing value count: 2,988

Missing value count: 0

Min	Mean	Median	Max	SD
0.08	0.19	0.19	0.29	0.03

<b>33</b>	<b>Column name:</b>	<b>pct_under_65_no_health_insurance</b>
-----------	---------------------	---

Column description: Percentage of population with no health insurance

Source information:		CHR		
Data type:		Numeric		
Unique non-missing value count:		2,986		
Missing value count:		0		
Min	Mean	Median	Max	SD
0.02	0.12	0.11	0.37	0.05

<b>34</b>	<b>Column name:</b>	<b>pct_highschool_diploma</b>		
Column description:		Percentage of adults ages 25 and over with a high school diploma or equivalent		
Source information:		CHR		
Data type:		Numeric		
Unique non-missing value count:		2,987		
Missing value count:		0		
Min	Mean	Median	Max	SD
0.50	0.88	0.89	0.99	0.06

<b>35</b>	<b>Column name:</b>	<b>inequality_ratio</b>		
Column description:		Ratio of household income at the 80th percentile to income at the 20th percentile		
Source information:		CHR		
Data type:		Numeric		
Unique non-missing value count:		2,988		
Missing value count:		0		
Min	Mean	Median	Max	SD
2.44	4.53	4.40	11.13	0.80

<b>36</b>	<b>Column name:</b>	<b>social_clubs_per_10k</b>		
Column description:		Number of membership associations per 10,000 population		
Source information:		CHR		

Data type:	Numeric			
Unique non-missing value count:	2,879			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.00	11.36	10.75	48.86	5.52

<b>37</b>	<b>Column name:</b>	<b>air_polution_metric</b>
-----------	---------------------	----------------------------

Column description:	Average daily density of fine particulate matter in micrograms per cubic meter			
Source information:	CHR			
Data type:	Numeric			
Unique non-missing value count:	103			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.90	7.59	7.80	15.60	1.68

<b>38</b>	<b>Column name:</b>	<b>water_quality</b>
-----------	---------------------	----------------------

Column description:	Indicator of the presence of health-related drinking water violations. 0=No, 1=Yes			
Source information:	CHR			
Data type:	Numeric			
Unique non-missing value count:	2			
Missing value count:	0			
Categories	Frequency	Cumulative Frequency	Percent	
0	1,969	1,969	65.90	
1	1,019	2,988	34.10	

<b>39</b>	<b>Column name:</b>	<b>pct_high_housing_costs</b>
-----------	---------------------	-------------------------------

Column description:	Percentage of households with severe cost burden – monthly housing costs (including utilities) exceed 50% of monthly income			
Source information:	CHR			

Data type:	Numeric			
Unique non-missing value count:	2,964			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.02	0.10	0.10	0.25	0.03

<b>40</b>	<b>Column name:</b>	<b>pct_overcrowded_hh</b>
-----------	---------------------	---------------------------

Column description:	Percentage of households with overcrowding – more than 1 person per room			
Source information:	CHR			
Data type:	Numeric			
Unique non-missing value count:	2,888			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.00	0.02	0.02	0.15	0.02

<b>41</b>	<b>Column name:</b>	<b>pct_30_min_plus_commute</b>
-----------	---------------------	--------------------------------

Column description:	Percentage of workers who commute in their car alone, the percentage that commute more than 30 minutes			
Source information:	CHR			
Data type:	Numeric			
Unique non-missing value count:	559			
Missing value count:	0			
Min	Mean	Median	Max	SD
0.00	0.33	0.32	0.78	0.13

<b>42</b>	<b>Column name:</b>	<b>school_funding_gap</b>
-----------	---------------------	---------------------------

Column description:	The average gap in dollars between actual and required spending per pupil among public school districts. Required spending is an estimate of dollars needed to achieve U.S. average test scores in each district			
Source information:	CHR			

Data type: Numeric

Unique non-missing value count: 2,988

Missing value count: 0

Min	Mean	Median	Max	SD
-18852.82	268.68	443.08	27719.24	4745.88

<b>43</b>	<b>Column name:</b>	<b>pct_voters</b>
-----------	---------------------	-------------------

Column description: Percentage for voter turnout

Source information: CHR

Data type: Numeric

Unique non-missing value count: 2,987

Missing value count: 0

Min	Mean	Median	Max	SD
0.19	0.65	0.66	1.00	0.10

<b>44</b>	<b>Column name:</b>	<b>pct_home_owner</b>
-----------	---------------------	-----------------------

Column description: Percentage of owner-occupied housing units

Source information: CHR

Data type: Numeric

Unique non-missing value count: 2,776

Missing value count: 0

Min	Mean	Median	Max	SD
130.00	25783.99	7389.50	1545929.00	68063.66

<b>45</b>	<b>Column name:</b>	<b>pct_65_plus</b>
-----------	---------------------	--------------------

Column description: Percentage of 65 and Older raw value

Source information: CHR

Data type: Numeric

Unique non-missing value count: 2,988

Missing value count:		0		
Min	Mean	Median	Max	SD
0.05	0.20	0.20	0.58	0.05

<b>46</b>	<b>Column name:</b>	<b>pct_rural_population</b>
-----------	---------------------	-----------------------------

Column description:		Percentage Rural raw value		
Source information:		CHR		
Data type:		Numeric		
Unique non-missing value count:		2,346		
Missing value count:		0		
Min	Mean	Median	Max	SD
0.00	0.59	0.59	1.00	0.31

<b>47</b>	<b>Column name:</b>	<b>response</b>
-----------	---------------------	-----------------

Column description:		Response variable, percentage of adults reporting fair or poor health		
Source information:		Calculated as Better or Worse		
Data type:		Factor		
Unique non-missing value count:		2		
Missing value count:		0		
Categories	Frequency	Cumulative Frequency	Percent	
better	1,477	1,477	49.43	
worse	1,511	2,988	50.57	