Mapping Social Determinants of Health in Utah: A Data-driven Approach for Public Health and Clinical Decision-making

Sunho Im • Joshua Kawasaki • Kristin Knippenberg • Mahony Reategui-Rivera • L Weaver

GitHub https://github.com/lwynholds/BMI6016_SDOH

Group 4, BMI 6016 Data Wrangling, Spring 2025

Department of Biomedical Informatics
University of Utah School of Medicine

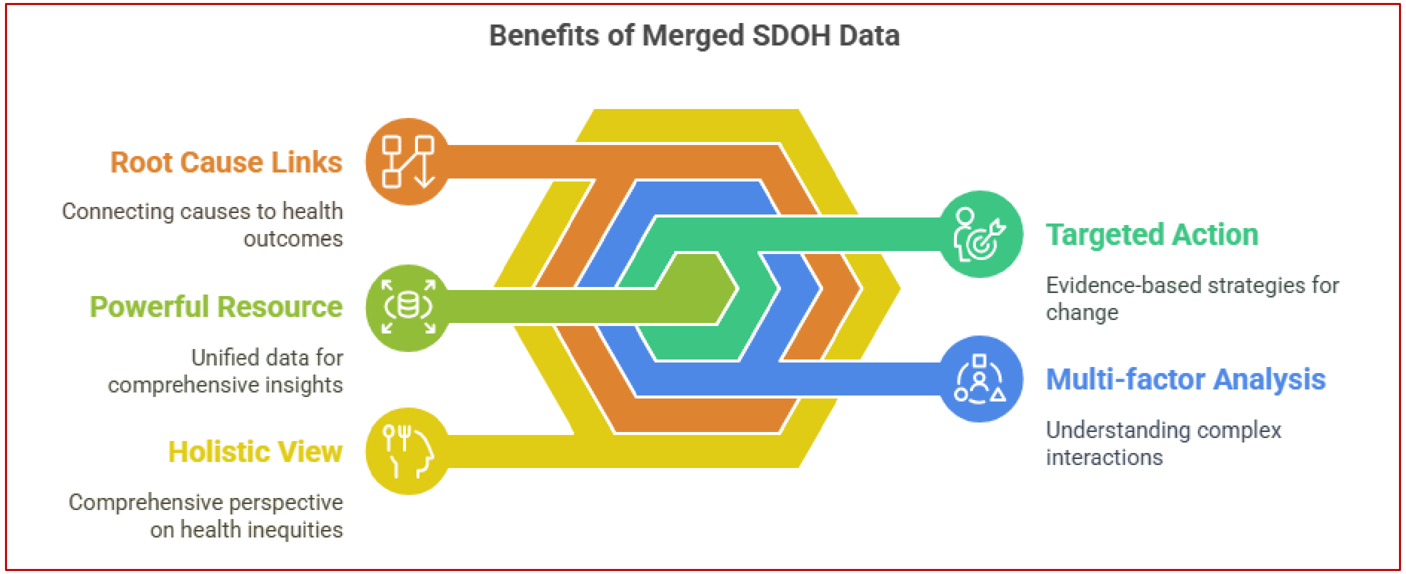
HEALTH
UNIVERSITY OF UTAH

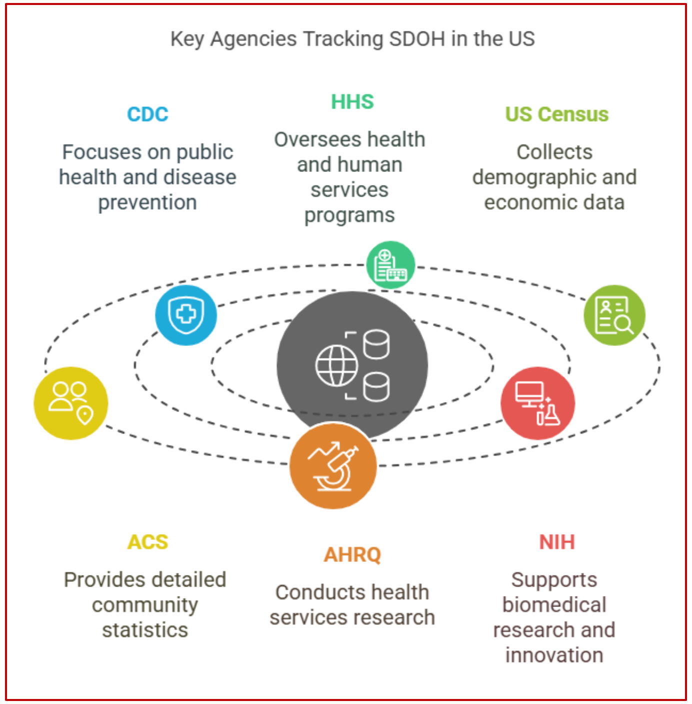
April 22, 2025

BACKGROUND AND MOTIVATION

- **Social Determinants of Health (SDOH)** are *"the circumstances in which people are born, grow up, live, work and age, and the systems put in place to deal with illness. These circumstances are in turn shaped by a wider set of forces: economics, social policies, and politics"*

- In 2018, Cantor et al. published their Factors Affecting Communities and Enabling Targeted Services (FACETS) open-data model
    - to standardize and compile SDOH-related data at the census-tract level in New York City, and
    - to integrate community-level determinants with patient health records in clinical settings





- **We propose to develop a FACETS-based model tailored to the state of Utah**, by integrating Utah-specific SDOH data, including environmental data, into a framework, we aim to provide a comprehensive tool for assessing community-level health determinants to support informed public health interventions and patient-level clinical decision-making

Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

HEALTH
UNIVERSITY OF UTAH

April 22, 2025

PROJECT OBJECTIVES

- To **standardize Utah's SDOH data at the US census-tract level** by integrating multiple sources of publicly available and state-specific data, using:
  - ACS and US Census Bureau, CDC, Environmental Protection Agency (EPA), and US Department of Agriculture Food Access Research Atlas (USDA) data (data sources from the original FACETS study that are also relevant to Utah)
  - State datasets from the Utah Department of Health and Human Services
  - State SDOH variables not otherwise specified that are particular or pertinent to Utah (e.g., urban/rural, electoral participation)
  - Temporal, categorical, or spatial crosswalks, where necessary

- To **create an interoperable dataset** of census tracts, within which an individual patient's address can be mapped to a Utah FACETS characterization of SDOH context

- To **characterize disparities in health outcomes** based on the intersection of air quality with other environmental health factors and SDOH.

- To **give context to any individual patient's health status** and assist patients and providers in deciding how best to improve that patient's health.

Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

DATASETS

Social Vulnerability Index
SVI percentile ranking •
English proficiency (2020)

American Community Survey
Census tract geodata •
Urban-Area classification
(2020)

$PM_{2.5}$ Max Value •
Respiratory hazard/cancer
risk (2020)

Asthma prevalence
(2022)

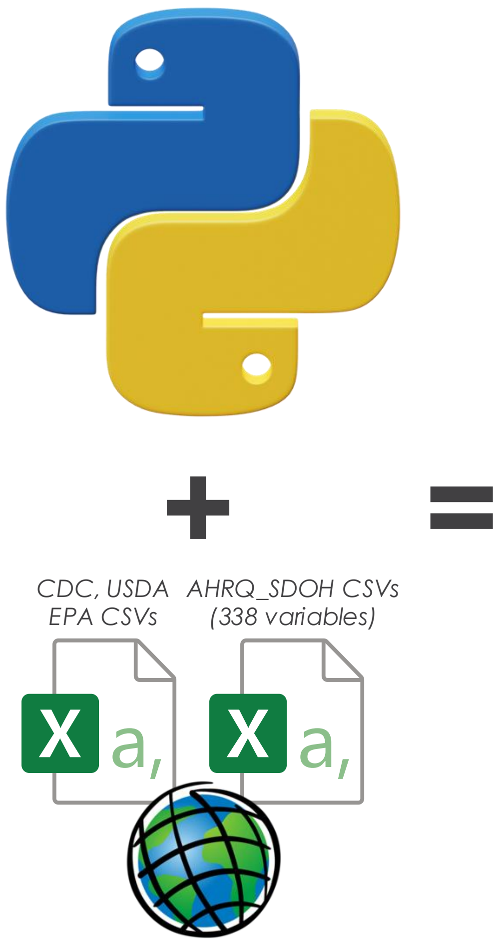
Agency for Healthcare Research and Quality

Total population • Citizenship • Racial/Ethnic diversity •
Unemployment • Health insurance status • Housing •
Education • Poverty • Household income • Walkability •
GINI Index of Inequality (2020)

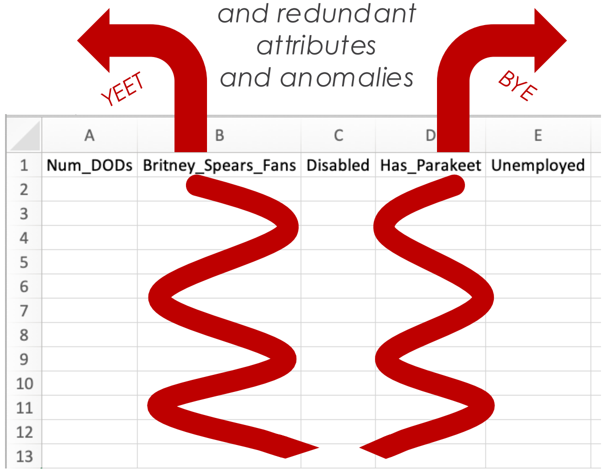
Law enforcement and
municipal geodata •
Crime (2022)

Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

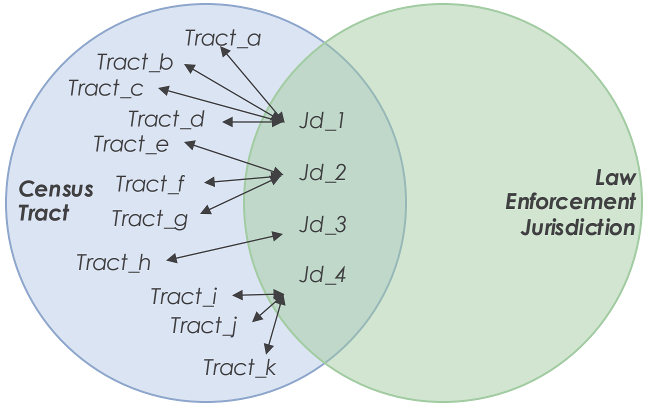
DATA PROCESSING AND WRANGLING



Remove unwanted and redundant attributes and anomalies



Turn jurisdictional and county data into census tracts with CROSSWALKS



Merge, join, and transform variables into domains, themes, and layers.

First a lot of this …

```python
# Slim down the df
svi_kept_var = ['ST', 'STATE', 'ST_ABBR', 'STCNTY', 'COUNTY', 'FIPS', 'LOCATION',
'AREA_SQMI', 'E_TOTPOP', 'RPL_THEMES', 'EP_LIMENG']
svi_dff = svi_df[svi_kept_var].copy()

#Fix the joining variable (it has 11 digits already; just needs to be a string
and renamed) and put it near the other variables
svi_dff['GEOID20'] = svi_dff['FIPS'].astype("string")
svi_dff = svi_dff[['ST', 'STATE', 'ST_ABBR', 'STCNTY', 'COUNTY', 'GEOID20'] +
[col for col in svi_dff.columns if col not in ['ST', 'STATE', 'ST_ABBR',
'STCNTY', 'COUNTY', 'GEOID20']]]
svi_dff.replace(-999, np.nan, inplace=True)

#svi_dff.head()
#svi_dff.info()
#print(svi_dff.isna().sum())
```

… with this as the goal …

AIR QUALITY	% with asthma	IDENTITY	Born citizens vs foreign-born
	Cancer risk per 1M		% % non-Hispanic, white
	PM2.5 Max	ECONOMIC ISSUES	Median household income
SOCIETY	% speaking limited English		% housing units owned, rented, or vacant
	Avg education level		% paying >30% for housing
	Crimes per 1K		% unemployed or no health insurance
	GINI index		% under the poverty line
	SVI percentile ranking	CIVIC	"Urban" areas and Walkability

Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

RESULTS: VARIABLE DESCRIPTIONS

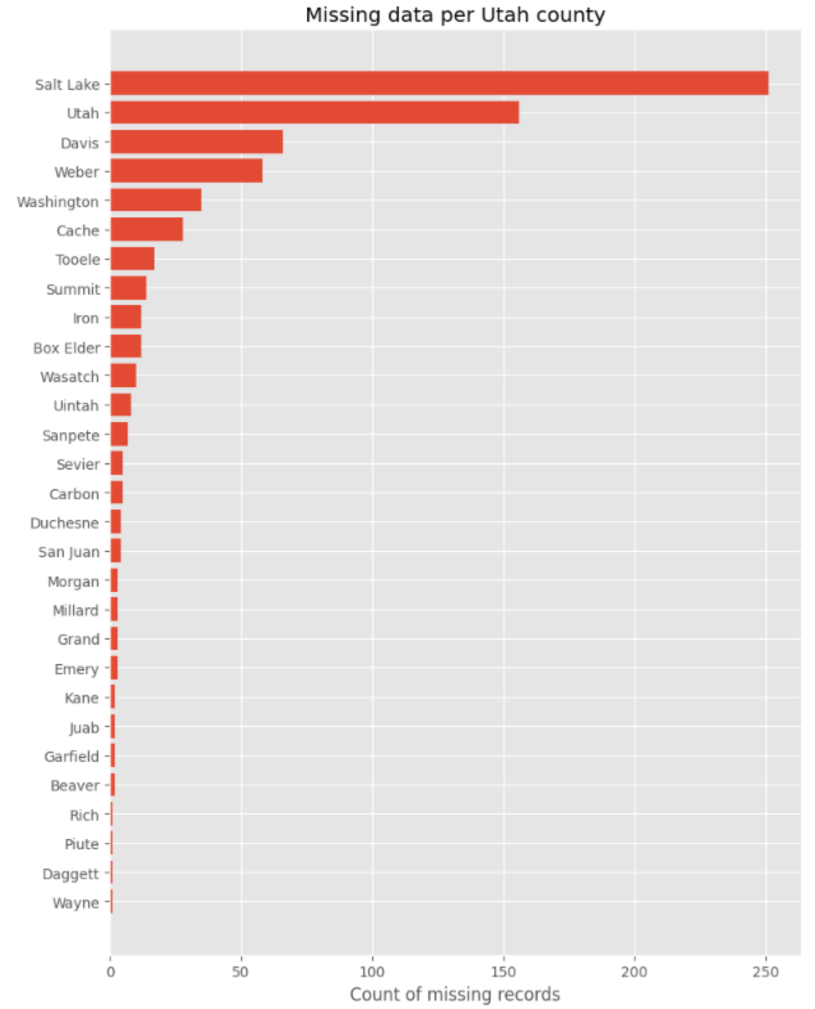
Dimensions of the DF: (716, 38)

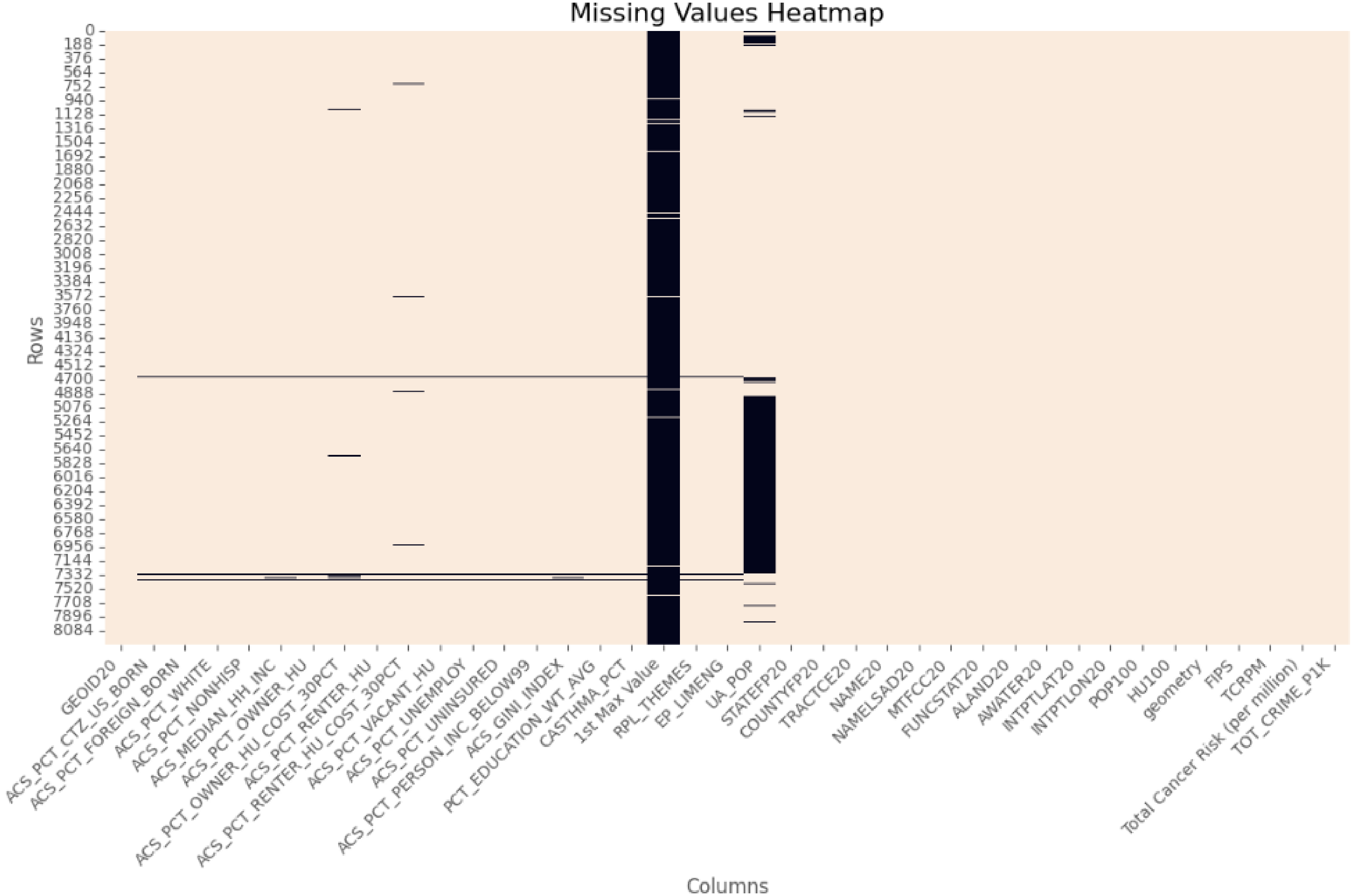
Variables	Descriptions	Variables (continued)	Descriptions (continued)
GEOID20	Census tract identifier (2020)	ACS_PCT_UNINSURED	% of population without health insurance
ACS_PCT_CTZ_US_BORN	% of U.S.-born citizens	ACS_PCT_PERSON_INC_BELOW99	% of people with income below the poverty line
ACS_PCT_FOREIGN_BORN	% of foreign-born residents	ACS_GINI_INDEX	Income inequality index
ACS_PCT_WHITE	% identifying as white	PCT_EDUCATION_WT_AVG	Weighted average of educational attainment
ACS_PCT_NONHISP	% non-Hispanic	CASTHMA_PCT	% of population with asthma
ACS_MEDIAN_HH_INC	Median household income	RPL_THEMES	CDC Social Vulnerability Index score
ACS_PCT_OWNER_HU	% owner-occupied housing units	EP_LIMENG	% of people with limited English proficiency
ACS_PCT_OWNER_HU_COST_30PCT	% owners with housing costs > 30% of income	UA_POP	Population in Census-designated "Urban Areas"
ACS_PCT_RENTER_HU	% renter-occupied housing units	ACS_PCT_WALK_2WORK	% of workers walking to work
ACS_PCT_RENTER_HU_COST_30PCT	% renters with housing costs > 30% of income	1st Max Value	PM25 highest µg/m³ at monitoring site
ACS_PCT_VACANT_HU	% vacant housing units	TCPRM	Total cancer risk (per million)
ACS_PCT_UNEMPLOY	% unemployed individuals	TOT_CRIME_P1K	Total crimes (per 1000)

Department of Biomedical Informatics
University of Utah School of Medicine

HEALTH
UNIVERSITY OF UTAH

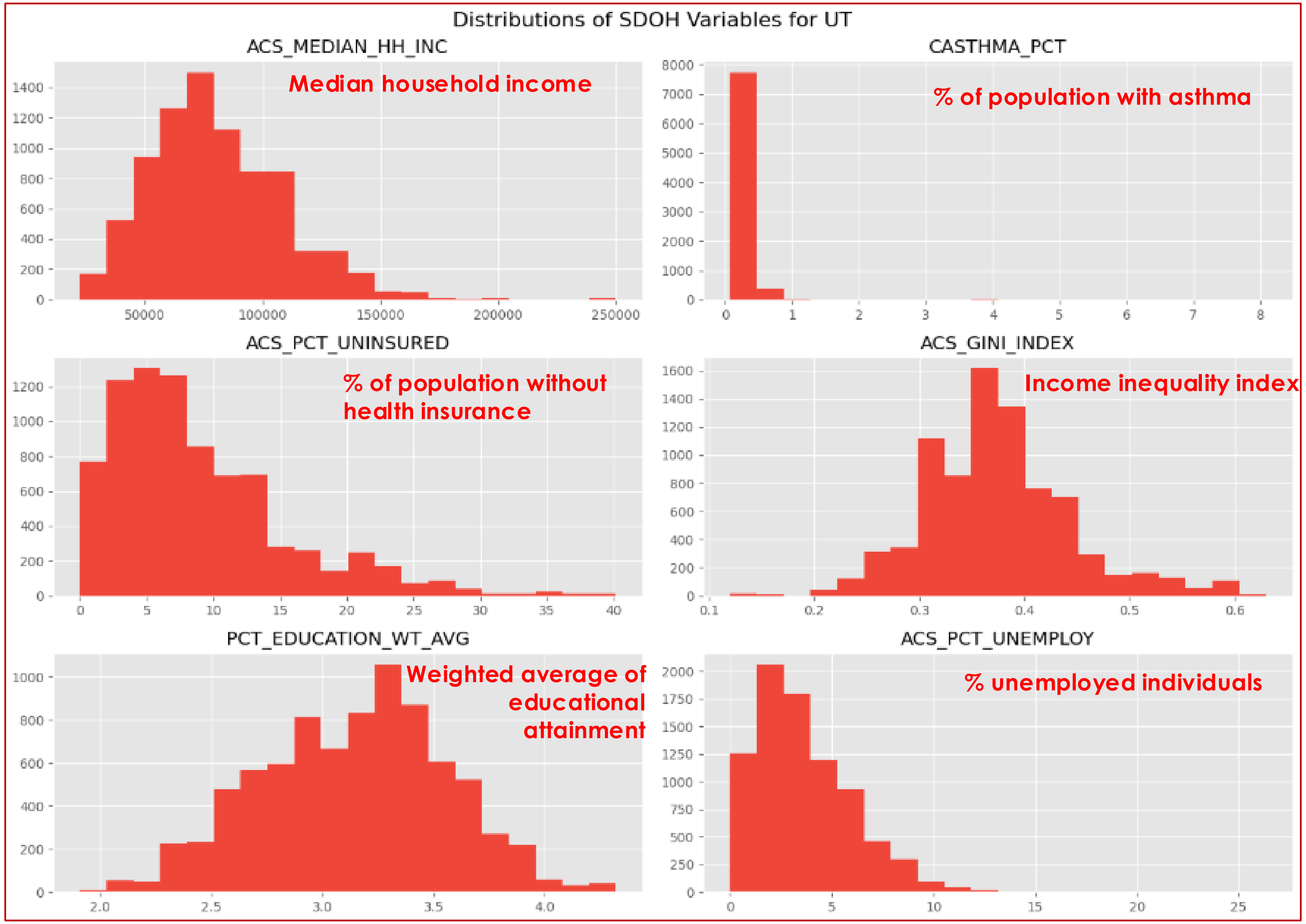
RESULTS: MISSING DATA



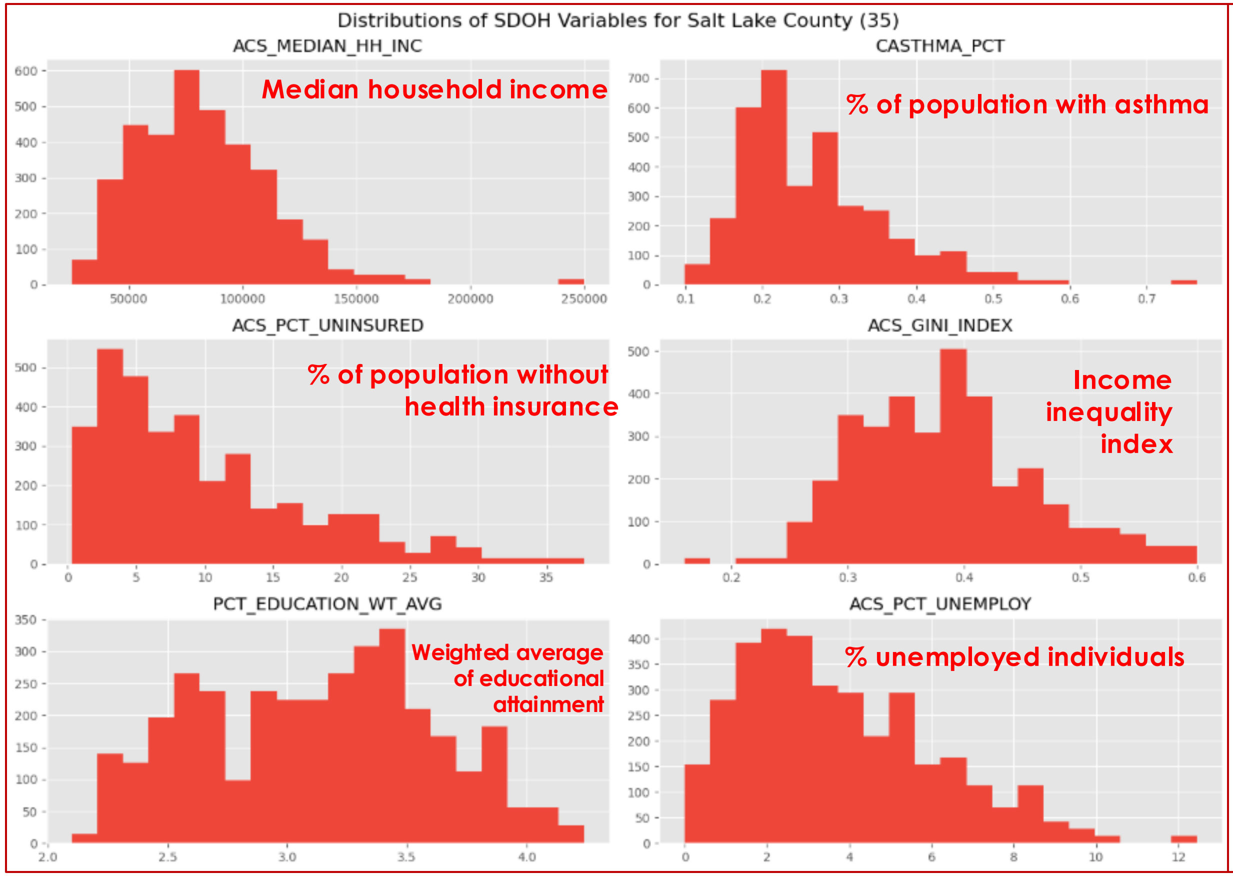


Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

RESULTS: EDA BY STATE LEVEL



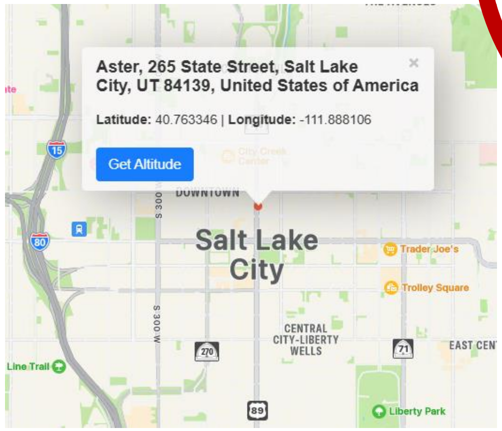
RESULTS: EDA BY COUNTY LEVEL



DATA ANALYSIS

```python
latitude_slc = 40.763346
longitude_slc = -111.888106

sdoh_result = get_sdoh_for_coordinates(latitude_slc, longitude_slc, sdoh_df=df)
print(sdoh_result)
```



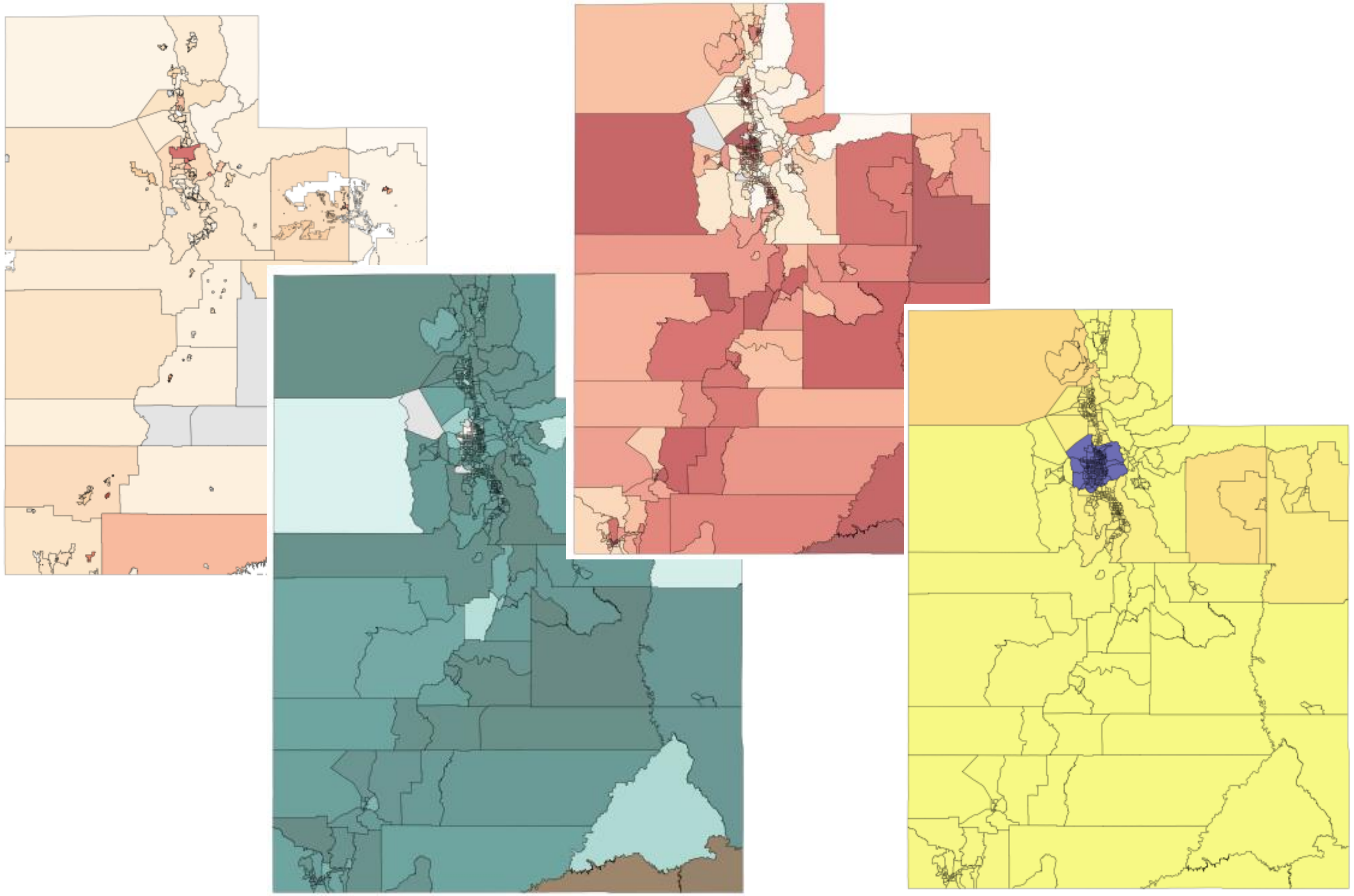

```
input_point                                          POINT (-111.888106 40.763346)
STATEFP20                                                                       49
COUNTYFP20                                                                     035
TRACTCE20                                                                   114000
GEOID20                                                                49035114000
NAME20                                                                        1140
NAMELSAD20                                                            Census Tract
MTFCC20                                                                      G5020
FUNCSTAT20                                                                       S
ALAND20                                                                    3153173
AWATER20                                                                         0
INTPTLAT20                                                             +40.7575950
INTPTLON20                                                            -111.8974571
POP100                                                                        4344
HU100                                                                         2929
ACS_PCT_CTZ_US_BORN                                                             85
ACS_PCT_FOREIGN_BORN                                                            15
ACS_PCT_WHITE                                                                   84
ACS_PCT_NONHISP                                                                 94
ACS_MEDIAN_HH_INC                                                            61917
ACS_PCT_OWNER_HU                                                                38
ACS_PCT_OWNER_HU_COST_30PCT                                                     23
ACS_PCT_RENTER_HU                                                               62
ACS_PCT_RENTER_HU_COST_30PCT                                                    42
ACS_PCT_VACANT_HU                                                               21
ACS_PCT_UNEMPLOY                                                                 6
ACS_PCT_UNINSURED                                                                8
ACS_PCT_PERSON_INC_BELOW99                                                      11
ACS_GINI_INDEX                                                                   1
PCT_EDUCATION_WT_AVG                                                             4
CASTHMA_PCT                                                                      0
1st Max Value                                                                  NaN
RPL_THEMES                                                                       1
EP_LIMENG                                                                        1
UA_POP                                                                        4344
geometry                            POLYGON ((-111.91397866107363 40.7606410737250...
FIPS                                                                         49035
TCRPM                                                                          700
Total Cancer Risk (per million)                                                700
TOT_CRIME_P1K                                                                 1010
```

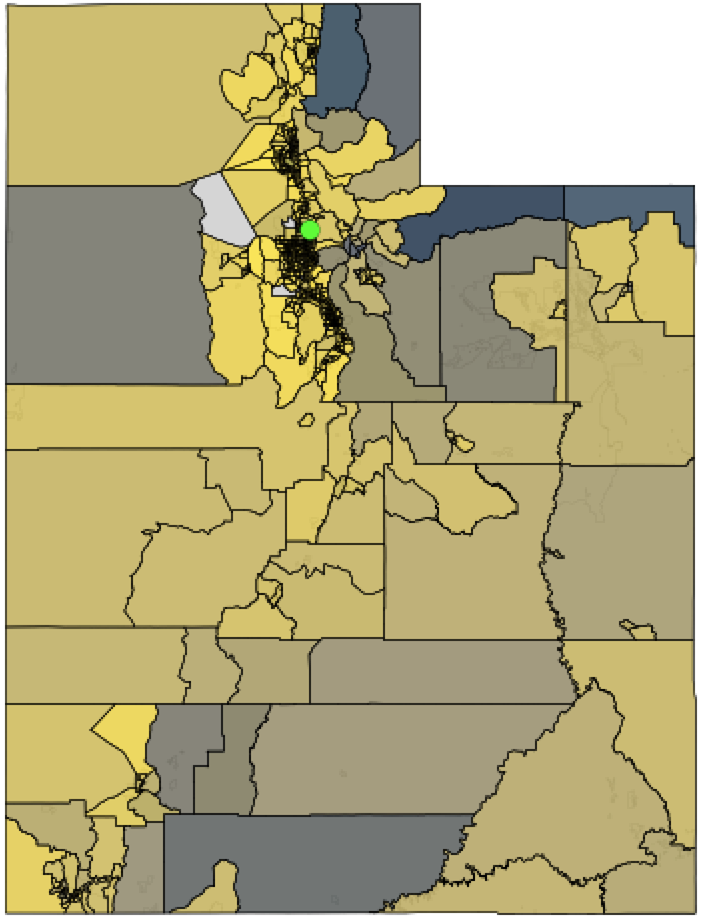
Inputting geographic coordinates within Utah allows us to identify the relevant census tract and retrieve the associated raw SDOH variables.

Here is an example for SLC

MAPS

By putting geospatial shape files into GeoDataFrames as basemaps, then linking each variable mostly by GEOID20 and assigning colormaps to the range of values, we made HEATMAPS:



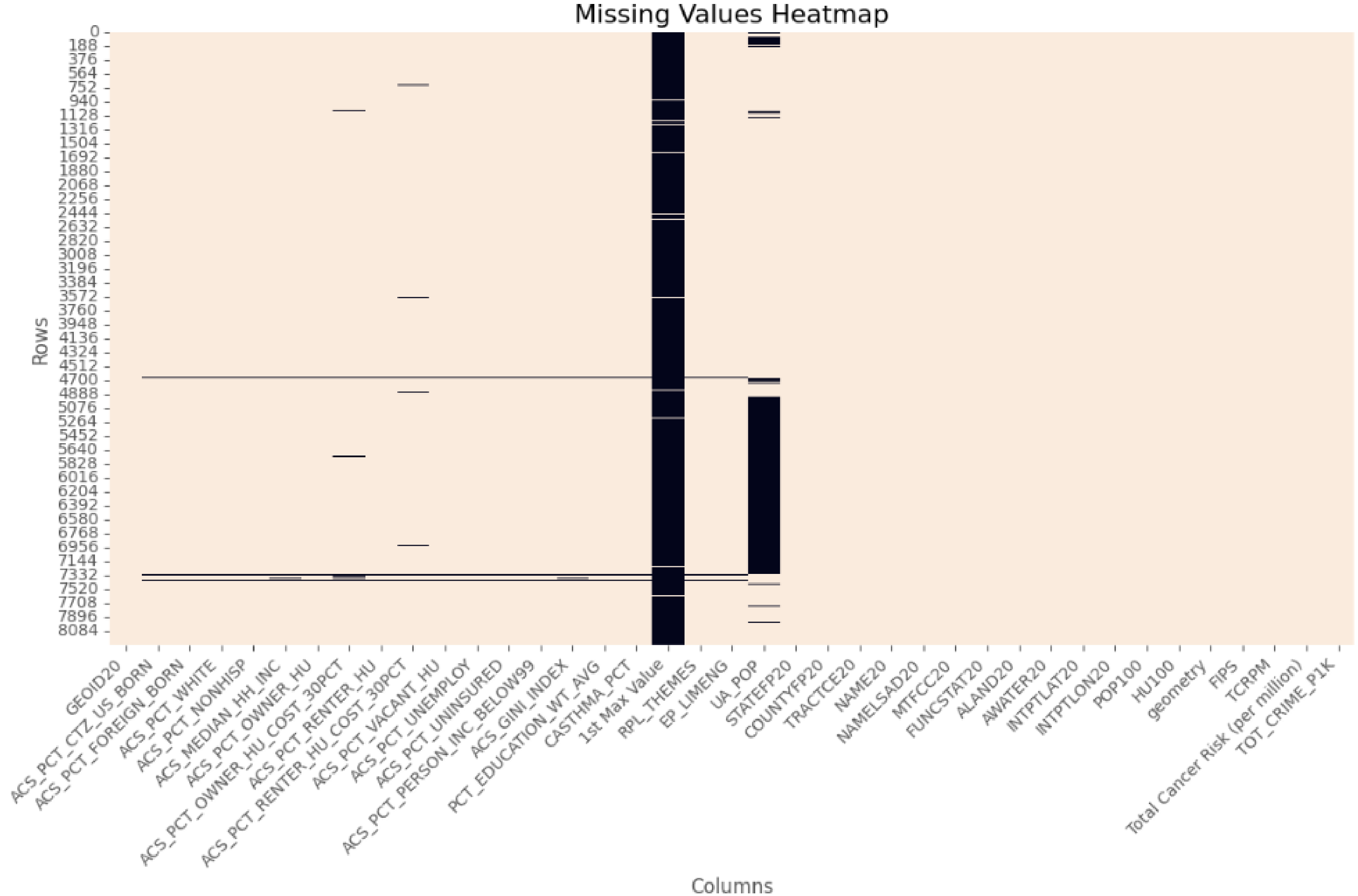


Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

CONCLUSIONS

- Data is readily available from public sources, but it requires varying degrees of data wrangling to compile them into one source that facilitates interoperability.
    - Mapping datasets that were originally compiled at the city or county level into census tracts leaves artifacts in the data
- This unified dataset, created through data compilation and mapping, enables aggregate analysis, GIS visualization, and comparisons of individual factor impact on clinical outcomes.
- This CSV dataset represents a significant first move towards standardizing and compiling SDOH data in an open architecture. This will allow for a better understanding of a patient's situation and improve decision-making in care planning.

POTHOLES IN THE INFORMATION SUPERHIGHWAY: MISSING DATA



Group 4, BMI 6016 Data Wrangling, Spring 2025
Department of Biomedical Informatics
University of Utah School of Medicine

REFERENCES

Social Determinants of Health (SDOH) Database (https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html)

Behavioral Risk Factor Surveillance System (https://www.cms.gov/About-CMS/Agency-Information/OMH/resource-center/hcps-and-researchers/data-tools/sgm-clearinghouse/brfss)

NaNDA (National Neighborhood Data Archive) (https://nanda.isr.umich.edu)

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: coming of age. Annu Rev Public Health, 32, 381-398. (https://doi.org/10.1146/annurev-publhealth-031210-101218)

Braveman, P., & Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. Public Health Rep, 129 Suppl 2(Suppl 2), 19-31. https://doi.org/10.1177/00333549141291S206

Canaday, F. T., Georas, S. N., & Croft, D. P. (2024). Examining the impact of air pollution, climate change, and social determinants of health on asthma and environmental justice. Curr Opin Pulm Med, 30(3), 276-280. (https://doi.org/10.1097/MCP.0000000000001065)

Cantor, M. N., Chandras, R., & Pulgarin, C. (2018). FACETS: using open data to measure community social determinants of health. J Am Med Inform Assoc, 25(4), 419-422. (https://doi.org/10.1093/jamia/ocx117)

Cook, Q., Argenio, K., & Lovinsky-Desir, S. (2021). The impact of environmental injustice and social determinants of health on the role of air pollution in asthma and allergic disease in the United States. J Allergy Clin Immunol, 148(5), 1089-1101 e1085. (https://doi.org/10.1016/j.jaci.2021.09.018)

Gao, C., Sanchez, K. M., & Lovinsky-Desir, S. (2023). Structural and Social Determinants of Inequitable Environmental Exposures in the United States. Clin Chest Med, 44(3), 451-467. (https://doi.org/10.1016/j.ccm.2023.03.002)

Marmot, M. (2005). Social determinants of health inequalities. Lancet, 365(9464), 1099-1104. (https://doi.org/10.1016/S0140-6736(05)71146-6)

Marmot, M., Allen, J., Bell, R., Bloomer, E., Goldblatt, P., Consortium for the European Review of Social Determinants of, H., & the Health, D. (2012). WHO European review of social determinants of health and the health divide. Lancet, 380(9846), 1011-1029. (https://doi.org/10.1016/S0140-6736(12)61228-8)

REFERENCES, CONTINUED

Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S., & Commission on Social Determinants of, H. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. Lancet, 372(9650), 1661-1669. https://doi.org/10.1016/S0140-6736(08)61690-6

McRoy, L., Rust, G., & Xu, J. (2017). Factors Associated with Asthma ED Visit Rates among Medicaid-enrolled Children: A Structural Equation Modeling Approach. AIMS Med Sci, 4(1), 71-82. https://doi.org/10.3934/medsci.2017.1.71

Office of Disease Prevention and Health Promotion. (2024, Jan 1). Social Determinants of Health Literature Summaries. US Department of Health and Human Services. Retrieved February 16 from https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries#block-sdohinfographics

The World Health Organization (WHO). (2003). Social Determinants of Health: The Solid Facts, 2nd edition. Retrieved from https://iris.who.int/handle/10665/326568

Zhu, G., Wen, Y., Cao, K., He, S., & Wang, T. (2024). A review of common statistical methods for dealing with multiple pollutant mixtures and multiple exposures. Front Public Health, 12, 1377685. https://doi.org/10.3389/fpubh.2024.1377685