

Template

Your original research question.

How do social determinants of health, such as income, education, and access to healthcare, affect health in underserved communities?

Steps used to find the dataset to attempt to answer your research question:

1. Step 1: I reviewed my research question and identified the health outcome (overall health outcomes) and social determinants (income, education, and healthcare access).
2. Step 2: I created keywords including “social determinants of health,” “health outcomes,” “income,” “education,” and “healthcare access.”
3. Step 3: I explored several public health data sources listed in the module.
4. ..Step 4: I compared datasets based on accessibility, coverage, and relevance to my research question.
5. ..Step 5: I selected the dataset that best matched both the health outcome and social determinant components of my question.

List the links to at least 2–3 datasets that you explored.

- Data.gov: www.data.gov
- World Health Organization (WHO): [The Global Health Observatory](http://www.who.int)
- HealthData.gov: [HealthData.gov](http://healthdata.gov).

Link to the dataset that you think is the most suitable.

<https://healthdata.gov>

Description of the dataset that you think is the most suitable. (about 50-80 words).

The HealthData.gov Social Determinants of Health dataset contains information related to healthcare access, education, income, economic stability, and community conditions across the United States. The data comes from public health agencies and government sources and covers multiple years and populations. It allows researchers to examine how social and environmental factors influence health outcomes and identify disparities among different communities.

Brief description of why you think this dataset can help you answer the research question.

This dataset is useful because it includes several social determinants of health that are central to my research question, such as income, education, and healthcare access. These

variables can be analyzed alongside health outcome measures to better understand health disparities in underserved communities.

How would you update/adapt the research question based on the data available?

How do income, education level, and healthcare access influence health outcomes among underserved communities in the United States?

Sample

Research Question: How does housing quality impact asthma rates in low-income neighborhoods?

Sources Explored:

Source	What I Found	Fit for My Question?
WHO Global Health Observatory	Global data on respiratory diseases; limited housing variables	Partial — has asthma data, not housing
HealthData.gov	U.S. federal health datasets; found CDC PLACES dataset	Good — has asthma prevalence by census tract
Kaggle	Community-uploaded datasets; found a merged housing + health dataset but source was unclear	Rejected — couldn't verify original source
Data.gov	Found HUD datasets on housing conditions in the U.S.	Good — has housing quality indicators

Final Dataset Selected: CDC PLACES: Local Data for Better Health (via [HealthData.gov](https://healthdata.gov/))

[Link]

- Source: Centers for Disease Control and Prevention (CDC), accessed through HealthData.gov
- Dataset Name: PLACES: Local Data for Better Health, Census Tract Data
- Data Type: Modeled estimates based on Behavioral Risk Factor Surveillance System (BRFSS) survey data
- Coverage: United States; census tract level; years 2020–2023
- Key Variables Available: Current asthma prevalence (%), poverty rates, access to health insurance, and other social determinants

Why I chose this:

It provides asthma rates at the census tract level, which means I can compare neighborhoods directly. It also includes socioeconomic indicators that relate to housing and income, making it very relevant to my research question.

After exploring the sources, I realized I have two paths forward:

Path 1: Stick with my original question, and pick the best available dataset

The CDC PLACES dataset is a strong fit and I can proceed with my original question as-is.

Final Dataset Selected: CDC PLACES: Local Data for Better Health (via HealthData.gov)

- **What's missing:** Direct housing quality measurements (e.g., mold, overcrowding). A follow-up dataset from HUD or the American Housing Survey could complement this.
-

Path 2: Tweak my question to better match what the data actually offers

During my exploration, I noticed that the WHO Global Health Observatory has strong global data on air pollution and respiratory disease — but it doesn't break down by neighborhood housing quality. This made me realize I could shift my question slightly to match what's more richly available.

Tweaked Research Question: *"How does household air pollution exposure impact asthma rates across low- and middle-income countries?"*

Why this tweak makes sense:

- "Housing quality" is hard to measure directly in most public datasets, but "household air pollution" (e.g., from cooking fuels) is a closely related and well-documented variable in WHO data.
- Shifting from "low-income neighborhoods" to "low- and middle-income countries" opens up more internationally comparable data.
- The core curiosity — the link between living conditions and respiratory health — stays the same.

Dataset for the Tweaked Question: WHO Global Health Observatory — Household Air Pollution Data

- **Source:** World Health Organization (WHO)

- **Data Type:** Country-level estimates compiled from national surveys and reports
 - **Coverage:** Global; 180+ countries; years 2010–2022
 - **Key Variables Available:** Percentage of population using clean fuels, age-standardized asthma mortality rates, country income classification
 - **Why this fits the tweaked question:** It directly connects fuel/air quality in homes with respiratory outcomes, across countries of varying income levels.
-

My Final Choice & Reasoning: I decided to go with **Path 1** and the CDC PLACES dataset because my original curiosity was specifically about neighborhood-level differences within a city, and this dataset preserves that granularity. However, I now know that if the data had been less available, tweaking my question to fit the WHO data would have been a completely valid and interesting direction too.

 **Key takeaway:** Your research question and your dataset should evolve together. It's okay — and actually good practice — to let the data inform how you refine your question.