

# Data Commons

Google | [datacommons.org](https://datacommons.org)

World Bank:  
GDP

NASA:  
Average  
Temperature

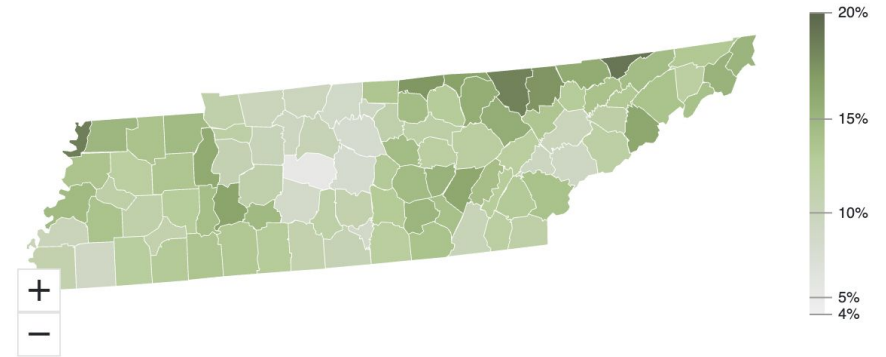


What Tennessee  
counties have the  
highest food insecurity?



Population: Food Insecure People in Counties of Tennessee (Per Capita in 2021)

Source: [feedingamerica.org](https://feedingamerica.org) • [Show metadata](#)



[Download](#) [Explore in Map Tool](#)

# Data is integral to nonprofit decision making

Proprietary + Confidential



## Strategic Planning

Where should we expand our services? Where are the most pressing needs?



## Fundraising & Grant Writing

How can we prove the need for our work?



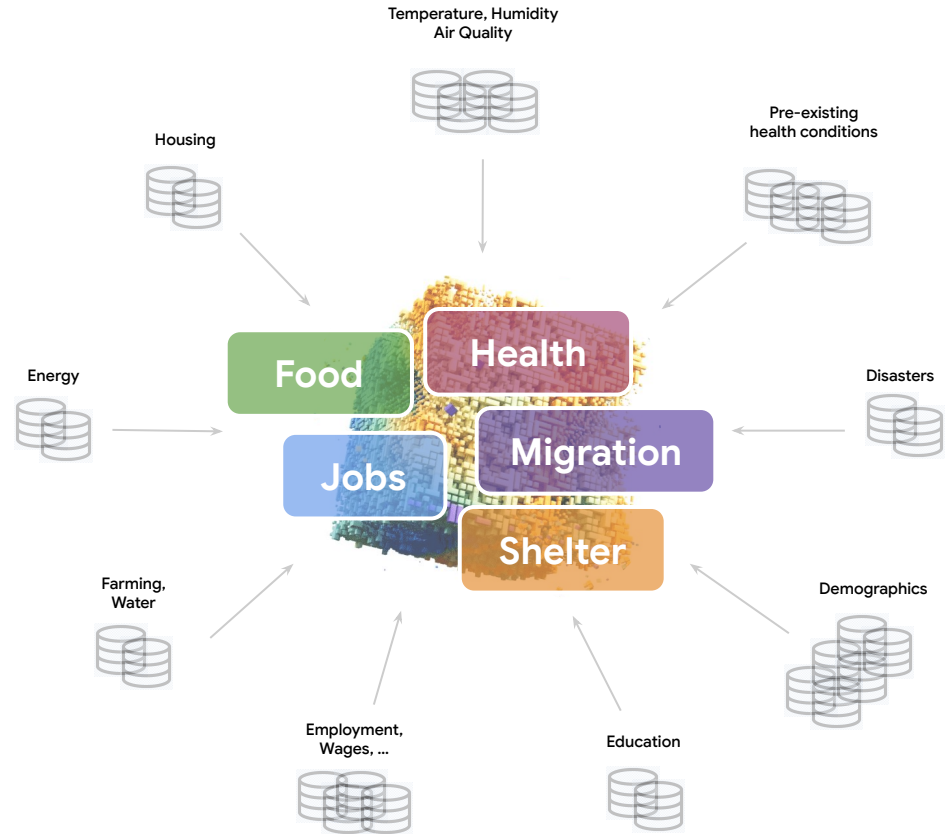
## Impact Measurement & Advocacy

Are our programs working?

# Why Data Commons

Proprietary + Confidential

- ✓ Many of the big challenges we face — climate change, increasing inequities, epidemics of diabetes — **will need deep, holistic insights** (data) to solve.
- ✓ **Complex challenges don't get solved with a single data source**
- ✓ **Data fragmentation** gets in the way of being used to make a big difference.



# About Data Commons

Proprietary + Confidential

**Mission** Organize all the world's public data and make it universally accessible and useful

---

**How** Open Source, Open Standards, Open APIs

A common knowledge graph:

Get data in schema.org formats

Using AI: Through natural language interface

---

**Progress** Data: 120+ sources, 250B+ data points

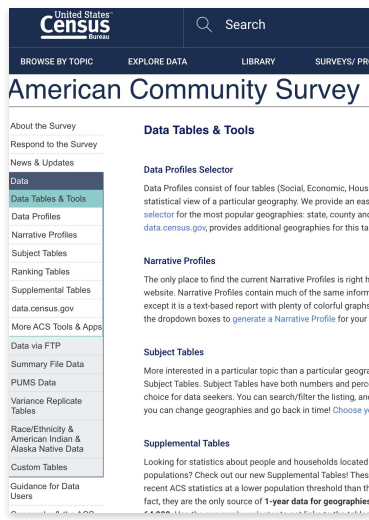
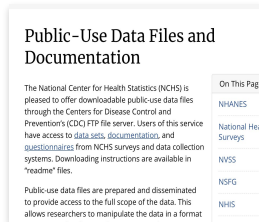
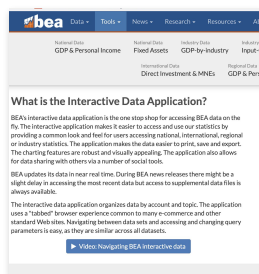
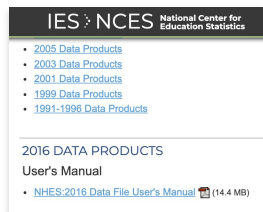
Partnerships: UN Stats, WHO, One.org, Stanford, etc.

The logo for Data Commons, consisting of the letters "DC" in white on a blue square background.

# Our goal for data

Proprietary + Confidential

From search for datasets,  
**Download, clean, normalize, join...**



Data Commons goal  
**Just ask in natural language**

Compare diabetes vs  
income in US counties



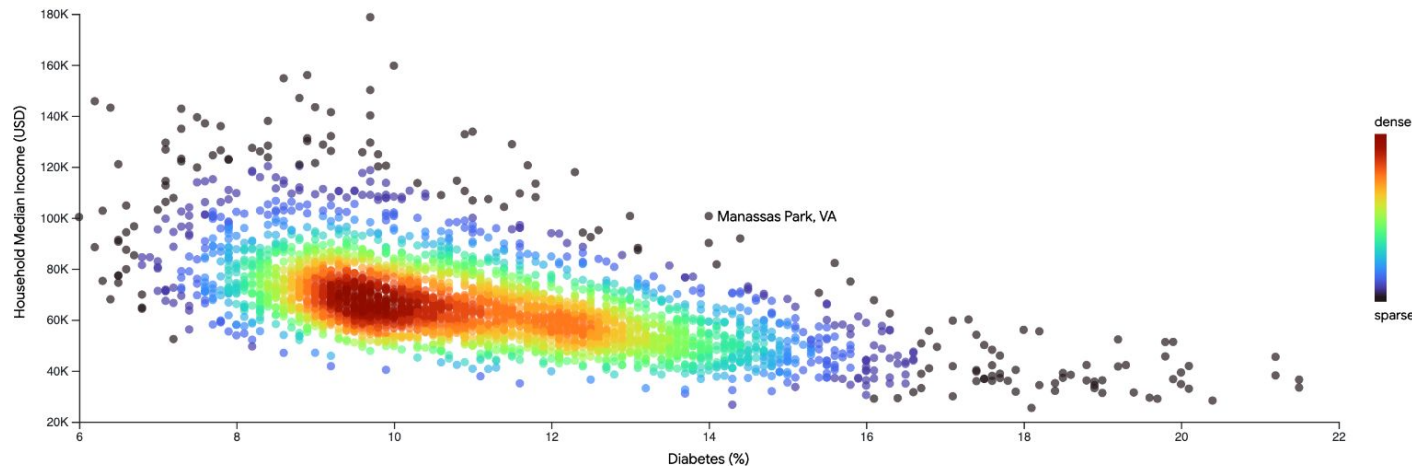
## United States • Income vs. Diabetes

Related topics [Individual Income Distribution](#) • [Individual Income by Industry](#) • [Household Income Distribution](#) • [Businesses](#) • [Economic Activity](#) • [Employment](#) • [Global Economic Activity](#) • [Poverty](#) • [Agriculture](#) • [Arthritis](#) • [Asthma](#) • [Blood Pressure](#) • [Cancer](#) • [Cholesterol](#) • [Heart Disease](#) • [Mental Health](#) • [Kidney Disease](#) • [Stroke](#) • [Heat Illnesses](#) • [Economy](#) • [Health Conditions](#)

### Household Median Income vs. Diabetes

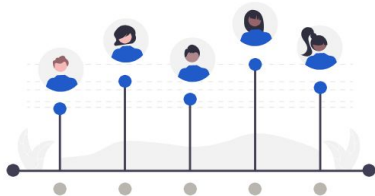
Household Median Income (2021 to 2023) Vs. Diabetes (2021 to 2022) in Counties of United States

Sources: [cdc.gov](#), [census.gov](#) • [Show metadata](#)





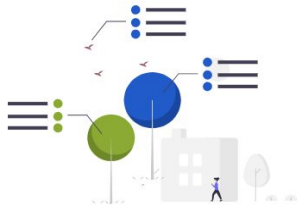
Economics



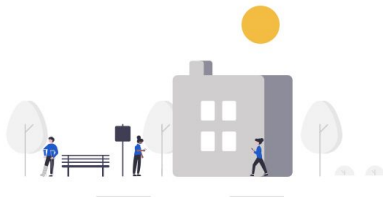
Demographics



Health



Sustainability



Education, Housing, and  
Commute



Equity



# Our partners

Proprietary + Confidential



UNITED NATIONS

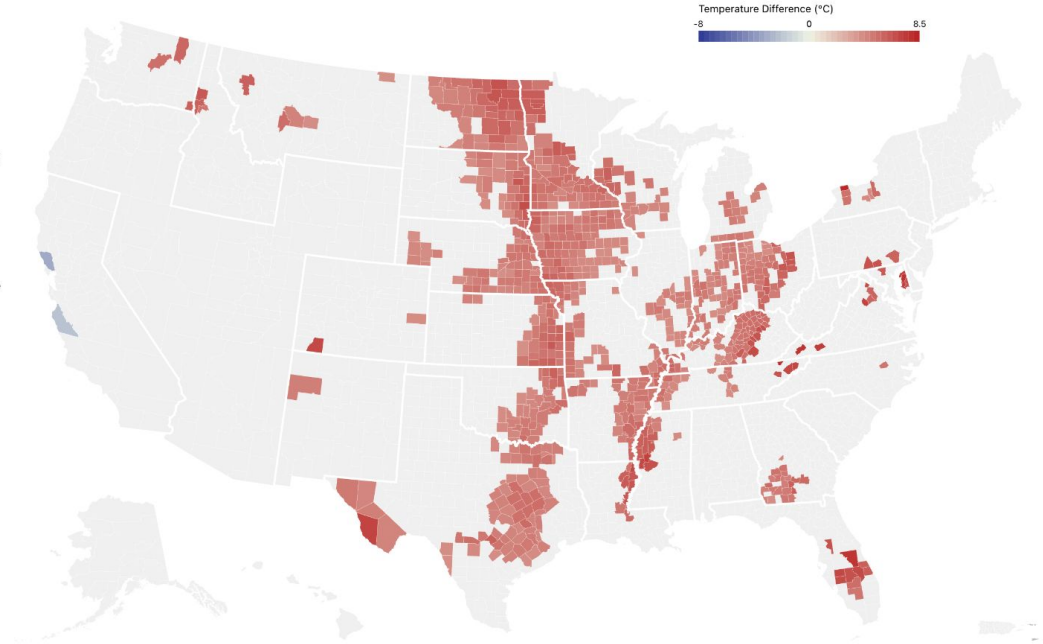


# Example: Techsoup and Feeding America

*To explore food security issues at a local level in Data Commons, click on any highlighted county on the map.*

To illustrate farming regions in the US which may be most impacted by climate change, the map now highlights counties with more than 50% farm cover that have extreme projected temperature differences (increase of more than 4 degrees or decrease of more than 2 degrees Celsius).

Click on a highlighted county to explore further in Data Commons.



# Build your own Custom Data Commons

Proprietary + Confidential

## Build and deploy your own

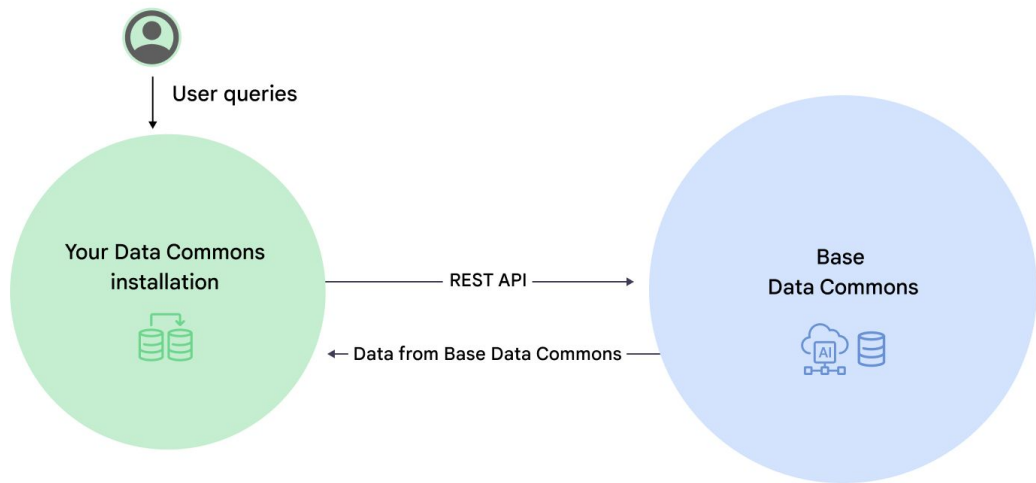
Launch your own Data Commons and customize it with your own data, with full control over your own data, computing resources, and access.

## Explore your data with natural language

Ask questions in your own words and get answers directly from your data.

## Gain new, actionable insights

Find actionable insights from your data in connection to global data.



# Data Gemma

Data Commons

## The opportunity • LLMs & accuracy

- Large Language Models (LLMs) struggle with real-world grounding, or provenance, lack clear sourcing, which can lead to hallucinations.

## The path to a solution • A future of grounded & trustworthy AI

- Can we use Data Commons to enhance LLM reasoning and accuracy?

## DataGemma are pioneering open models that help ground LLMs in real-world data through Data Commons

### Harnesses the knowledge of Data Commons

Leverages Data Commons' rich and ever-expanding knowledge resource and its natural language capabilities towards building more trustable AI.

### Enhances factuality and reasoning

Utilizes innovative retrieval techniques to provide factual and timely responses, overcoming the limitations of traditional LLMs.

### Responsible AI

Google's is committed to responsible AI for everyone and invests efforts into developing AI that is accurate, fair, and beneficial to society.

## What progress has Pakistan made against health goals?

### RAG ANSWER • EXAMPLE

Pakistan has made progress in some areas of public health, while others have stagnated or worsened. Positive trends are seen in:

- Life Expectancy: Increased from 62.102 years in 2000 to 66.098 years in 2021. [\[Table 5\]](#) This signifies improvement in overall health and a decrease in mortality rates.
- Access to Safe Drinking Water: Increased from 34.827% in 2000 to 50.6018% in 2022. [\[Table 1\]](#) This indicates progress in providing this essential service, although a significant portion of the population still lacks access.
- Hepatitis B Immunization Coverage: Increased significantly from 62% in 2003 to 85% in 2022. [\[Table 2\]](#) This demonstrates successful public health interventions targeting a specific disease.

...

[\[Table 1\]](#) • Proportion of Population Using Safely Managed Drinking Water Services, by Urban/Rural (%) in Pakistan (unit: Percentage), according to Global SDG Database

[\[Table 2\]](#) • Hepatitis B (Hepb3) Immunization Coverage Among 1-Year-Olds (%) in Pakistan, according to who.int

[\[Table 5\]](#) • Life Expectancy in Pakistan (unit: yr), according to datacatalog.worldbank.org

## Do the US states with high coal fired power also have high rates of COPD?

### RAG ANSWER • EXAMPLE

The provided data does suggest a correlation between coal consumption for electricity generation and rates of Chronic Obstructive Pulmonary Disease (COPD). Looking at the states with the highest coal consumption in 2020 [\[Table 2\]](#):

- Texas consumed 866,716,970 MMBtu of coal. They had a COPD rate of 6.2384% in 2021 [\[Table 4\]](#).
- West Virginia consumed 511,906,210 MMBtu of coal. They had a COPD rate of 9.6571% in 2021 [\[Table 4\]](#), the highest rate in the country.
- Indiana consumed 521,690,510 MMBtu of coal. They had a COPD rate of 7.505% in 2021 [\[Table 4\]](#).

Conversely, states with the lowest coal consumption in 2020 [\[Table 2\]](#) tended to have lower COPD rates:

- California consumed 1,322,520 MMBtu of coal. They had a COPD rate of 4.9821% in 2021 [\[Table 4\]](#).
- Connecticut consumed 66,720 MMBtu of coal. They had a COPD rate of 5.3503% in 2021 [\[Table 4\]](#).

...

**[Table 2]** • Consumption for Electricity Generation, Coal, All Sectors, Annual in States of United States (2020) (unit: MMBtu), according to eia.gov

**[Table 4]** • Chronic Obstructive Pulmonary Disease in States of United States (2020 to 2021) (unit: %), according to cdc.gov



# Thank you

If you had the world's public data at your fingertips, what would you ask?

Visit [datacommons.org](https://datacommons.org) to explore