

Data Democratization in the Age of AI

Oliver Wise, DOC Chief Data Officer



Office of the Under Secretary for
Economic Affairs



Open Data and AI: Putting it in Context



Questions on AI and Government

How can we promote responsible AI in society?

How do we leverage Gen AI within government?

How does government as a data producer leverage AI to democratize access to data?



History is likely to repeat itself

1. Technology emerges
2. Public expectations change
3. Policy follows

Credit: David Eaves



Type writers -> Paper size standards





Photocopier -> FOIA Laws



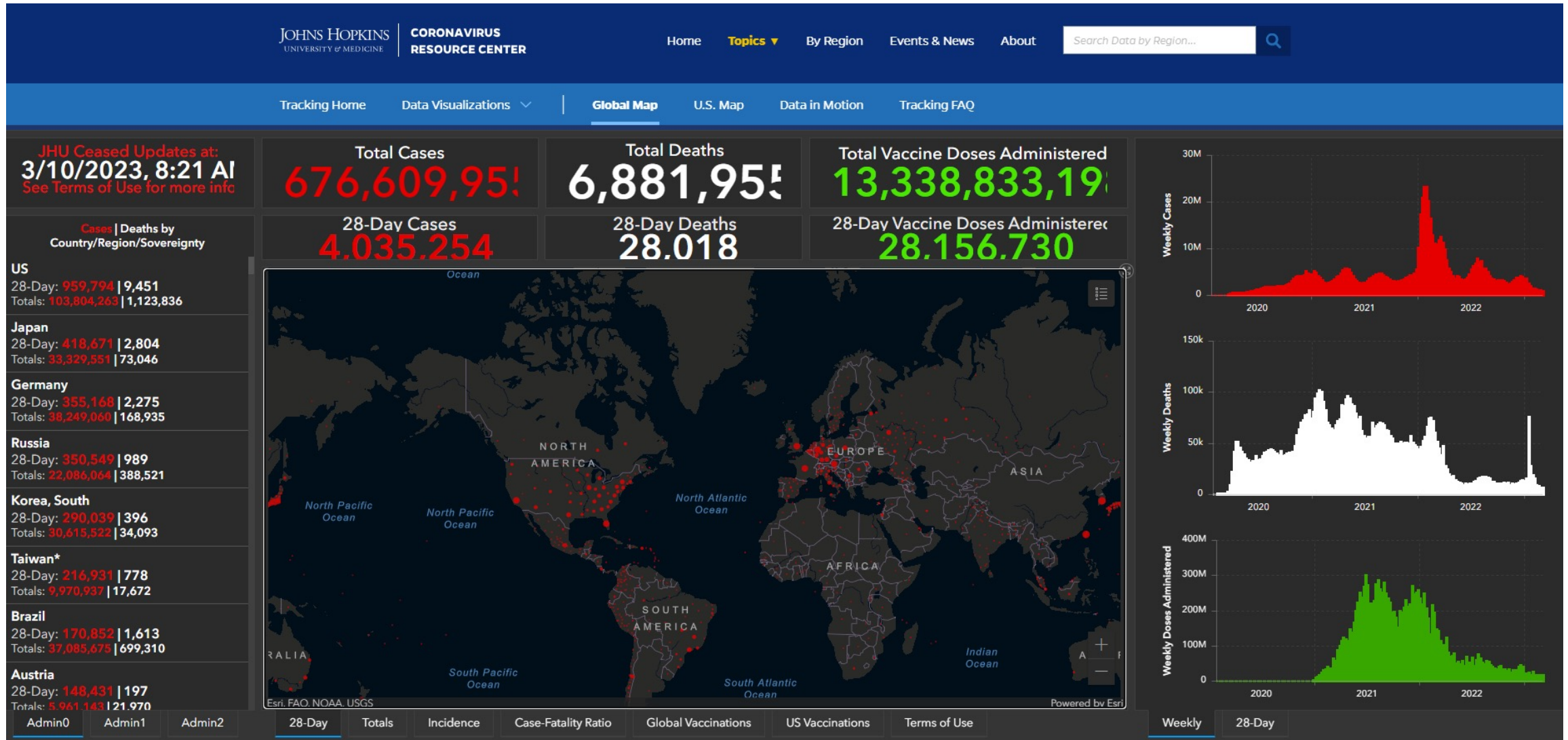


API -> Open Data EO & Data.gov



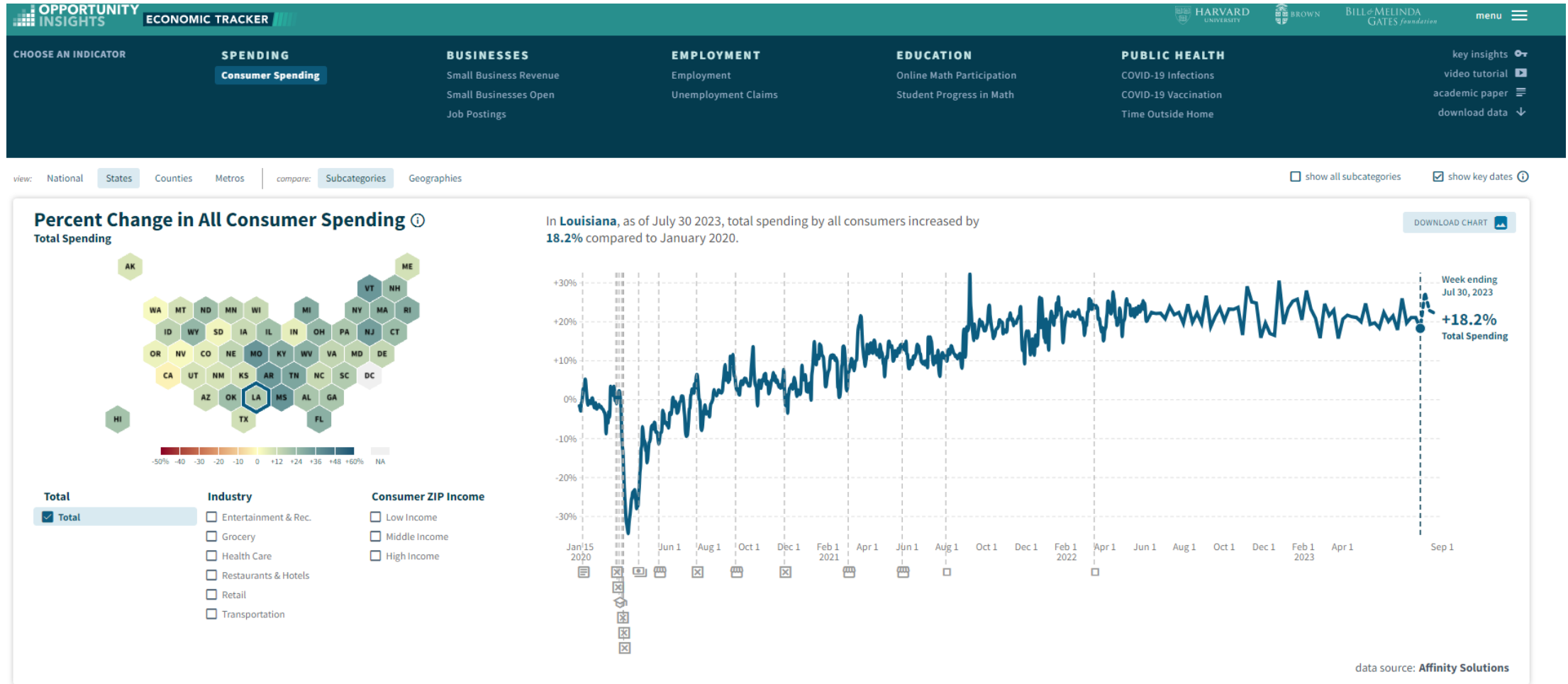


Open data in the pandemic: public health





Open Data in the pandemic: economic recovery





AI-Ready Open Government Data



AI has the power to vastly democratize insights from public data

- Simple, natural language interface
- Compelling, meaningful data visualizations
- Reliable, accurate results

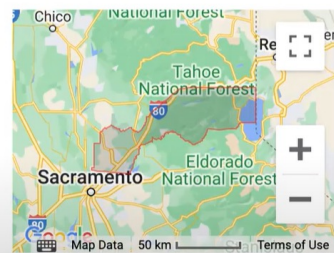
Data Commons NL Explorer

Explore Documentation About

Q: tell me more about Placer County

Feedback

Placer County



Rankings (in)	California	United States of America
Largest Population	22 of 59	176 of 3108
Highest Median Income	7 of 59	78 of 3187
Highest Median Age	18 of 59	1375 of 3187
Highest Unemployment Rate	49 of 59	1656 of 3185

Data from [census.gov](#), [fbi.gov](#), [bls.gov](#)

Learn more about Placer County: Economics, Health, Equity, Education, Demographics, Housing, Environment

What w

Beta [demo](#) of Google's Public Data Commons with natural language processing interface. Shared with permission.




However, we're not off to a great start

ChatGPT

zw Tell me about the demographics of suitland md

< 2 / 2 >

 Suitland is an unincorporated community in Prince George's County, Maryland, United States. According to the United States Census Bureau, the Suitland census-designated place (CDP) has a total area of 4.2 square miles (10.8 km²), all of it land. As of the 2020 census, the population of the Suitland CDP was 27,568.

In terms of demographics, the population of Suitland is predominantly African American, with African Americans comprising over 95% of the population. The remaining 5% is made up of people of other races, including White, Hispanic/Latino, Asian, and Native American. The median age in Suitland is 37.4 years, which is slightly younger than the median age of Maryland as a whole.

In terms of income, the median household income in Suitland is \$60,620, which is lower than the median household income for Prince George's County as a whole, but higher than the median household income for the state of Maryland. The poverty rate in Suitland is 13.9%, which is higher than the national average but lower than the poverty rate for Prince George's County.

Handwritten notes:
true (under 10.8 km²)
wrong (under 95%)
wrong² (under 37.4 years)
wrong (under 13.9%)
true (under lower than the poverty rate)

wrong

wrong²


wrong

wrong

true
wrong
true

Bard

Z Tell me about suitland md demographics

 View other drafts

Suitland is a town in Prince George's County, Maryland, United States. The population was 26,375 at the 2020 census. Suitland is part of the Washington metropolitan area.

The town is named after the Suitland Federal Research Center, which was established in 1939. The center is home to a number of federal agencies, including the National Archives and Records Administration, the United States Department of Agriculture, and the United States Department of Commerce.

Suitland is a diverse community with a population that is 89% African American, 6% Hispanic, and 2% white. The median household income is \$64,652.

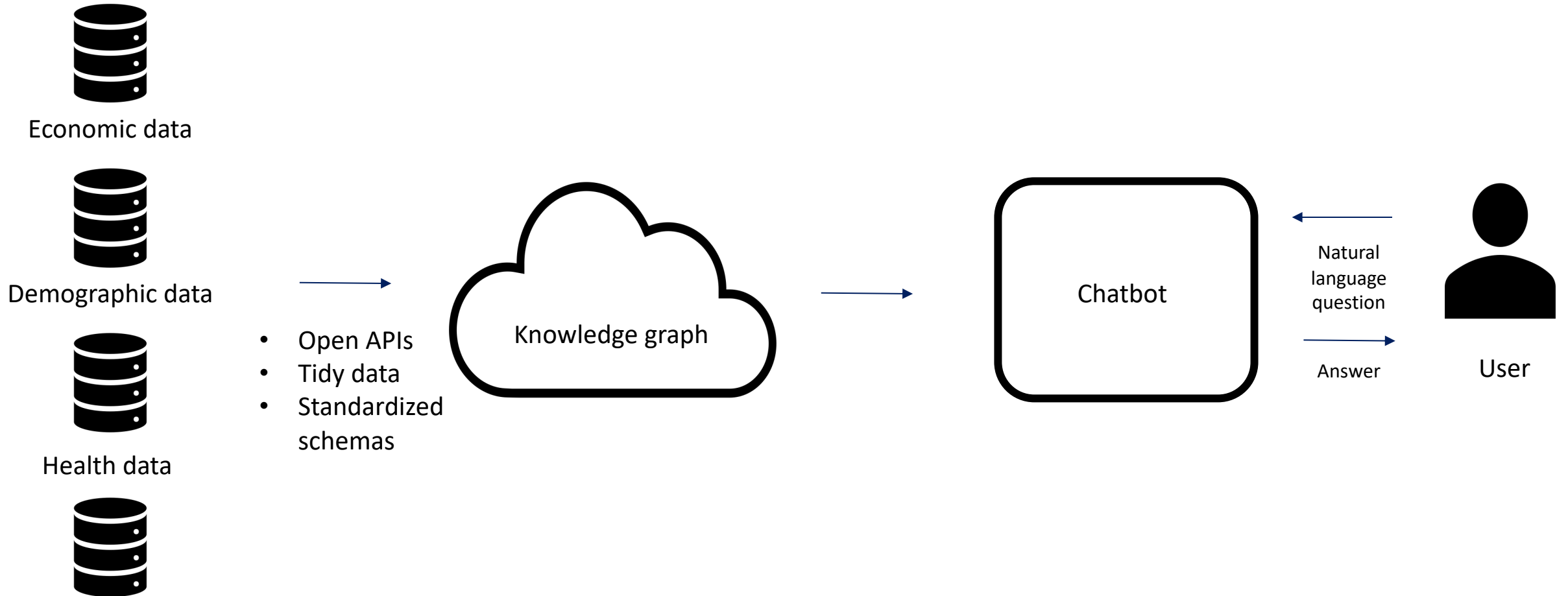
The town is home to a number of parks and recreation areas, including Suitland Park, which is home to a swimming pool, a playground, and a picnic area. The town is also home to the Suitland Library, which offers a variety of services for residents of all ages.

Suitland is a convenient location for commuters to Washington, D.C. The town is served by the

wrong



How do we make this work?



Credit: Swiss Federal Statistical Office

The background is a dark blue gradient with a faint, abstract network pattern of white lines and dots, resembling a molecular or data structure.

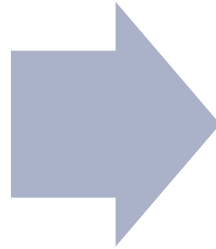
**We are not AI-Ready until our
data is.**



The data management leap we have to make

Machine readable

- Can be automatically processed by a computer
- Common formats (.csv, JSON, HTML)



Machine interpretable

- Tidy data
- Standardized schemas
- Rich metadata



Change in our APIs

Current ACS API

Name	Label
B01001A_001E	Estimate!!Total:
B01001A_001EA	Annotation of Estimate!!Total:
B01001A_001M	Margin of Error!!Total:
B01001A_001MA	Annotation of Margin of Error!!Total:
B01001A_002E	Estimate!!Total:!!Male:
B01001A_002EA	Annotation of Estimate!!Total:!!Male:
B01001A_002M	Margin of Error!!Total:!!Male:
B01001A_002MA	Annotation of Margin of Error!!Total:!!Male:
B01001A_003E	Estimate!!Total:!!Male:!!Under 5 years
B01001A_003EA	Annotation of Estimate!!Total:!!Male:!!Under 5 years
B01001A_003M	Margin of Error!!Total:!!Male:!!Under 5 Years

- 124 variables (1,240 for the 10 race/ethnicity-iterated tables)
- Complex variable names & labels

AI-Ready Data Example API

Name	Label
AGE_RANGE_CODE	Age range values
AGE_RANGE_LABEL	Description of Age range values
ESTIMATE	Population Estimate
GEO_ID	Geography
MOE	Margin of Error
NAME	Geographic Area Name
OLD_ITEM	Old Aggregate item name
RACE_CODE	Race Code
RACE_LABEL	Description of Race values
SEX_CODE	Sex Code
SEX_LABEL	Description of Sex values

- 11 variables (same 11 used across similar tables)
- Easy to understand, easy to use
- Additional Functionalities



...

1 to 30 of 30 |< < Page 1 of 1 > >



More problems to solve

- Capacity problem
- Fitness for purpose problem
- Data provenance problem
- Probabilistic model problem
- Velocity problem



Why does Open Data Matter?



Open data is foundational to our democracy

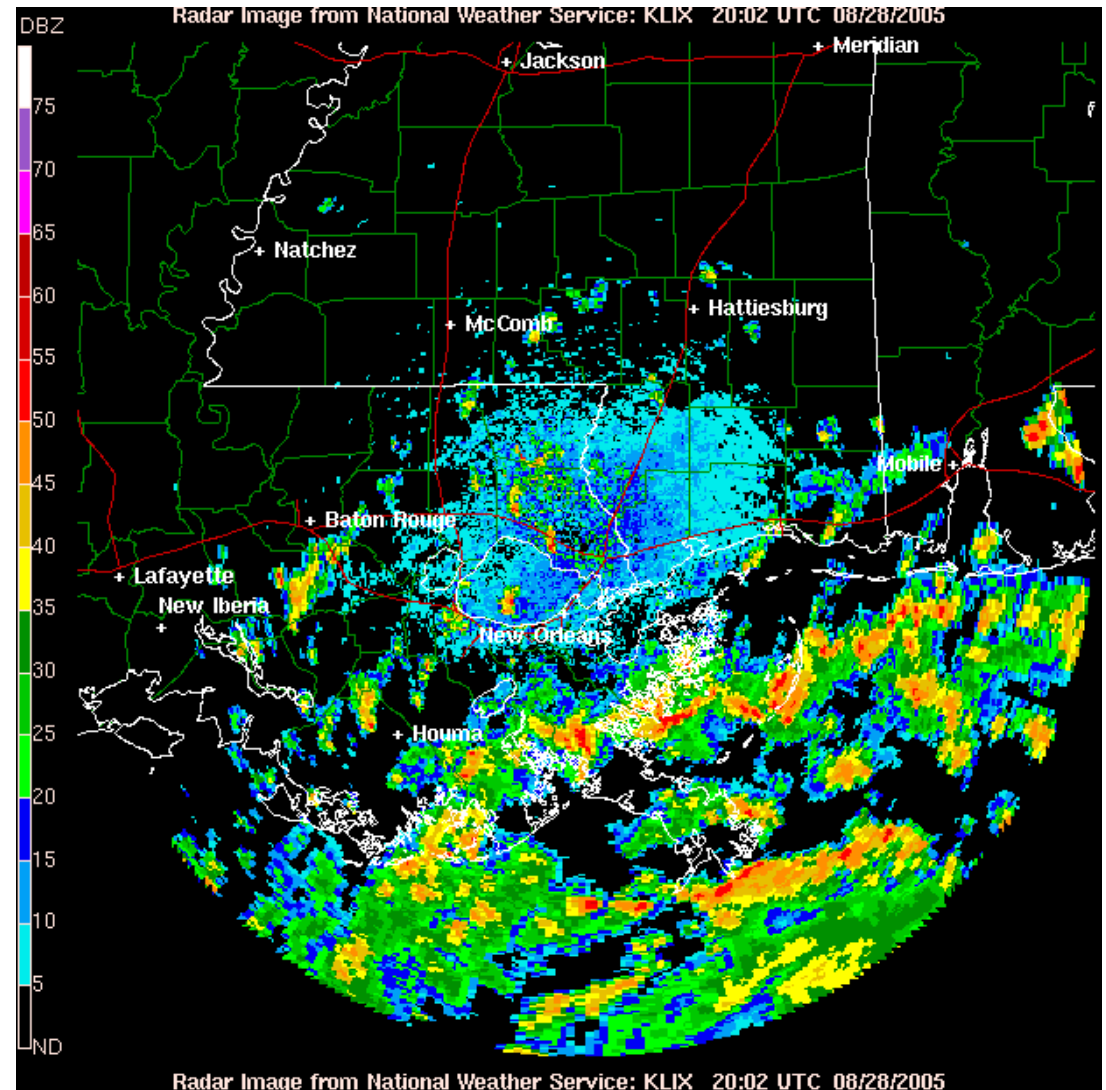




Open data drives scientific discovery

AI is increasingly integrated into scientific discovery to augment and accelerate research.

It enables scientists to gain a deeper understanding of complex systems and phenomena, leading to new discoveries and insights.

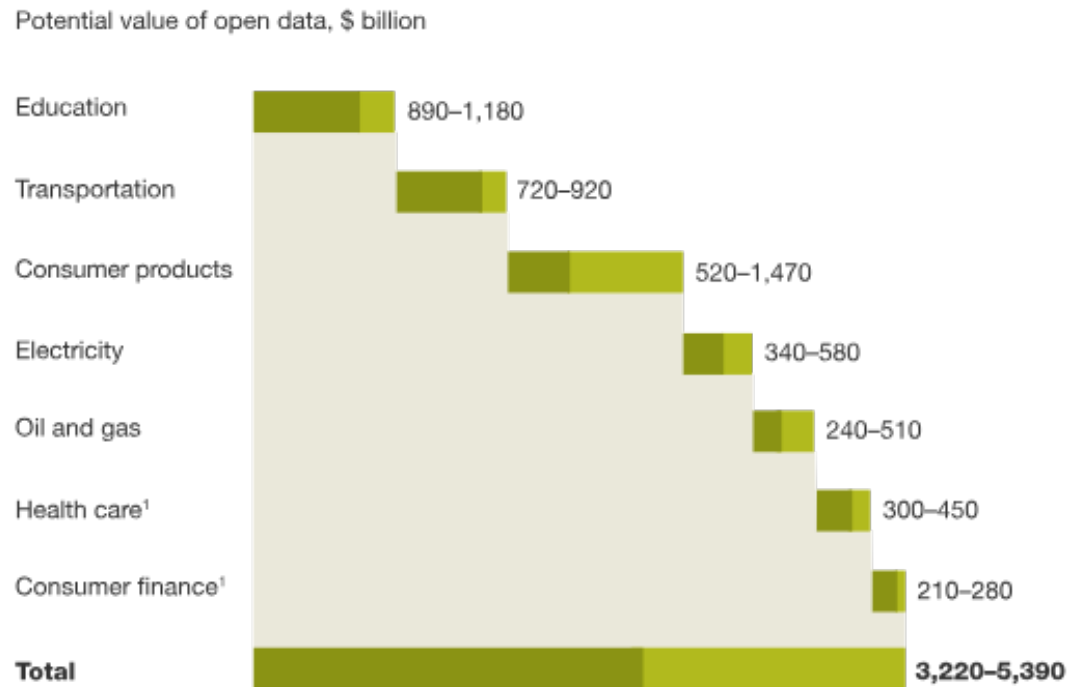


National Weather Service Radar during Katrina (NOAA)



Open data is key to driving economic growth

It is estimated that open data can help unlock **\$3 trillion to \$5 trillion in economic value annually** across several sectors.



¹Values for health care and consumer finance are drawn from examples of open-data potential and are not a comprehensive sizing of potential value across the 2 sectors; health-care figures are based on US values only.

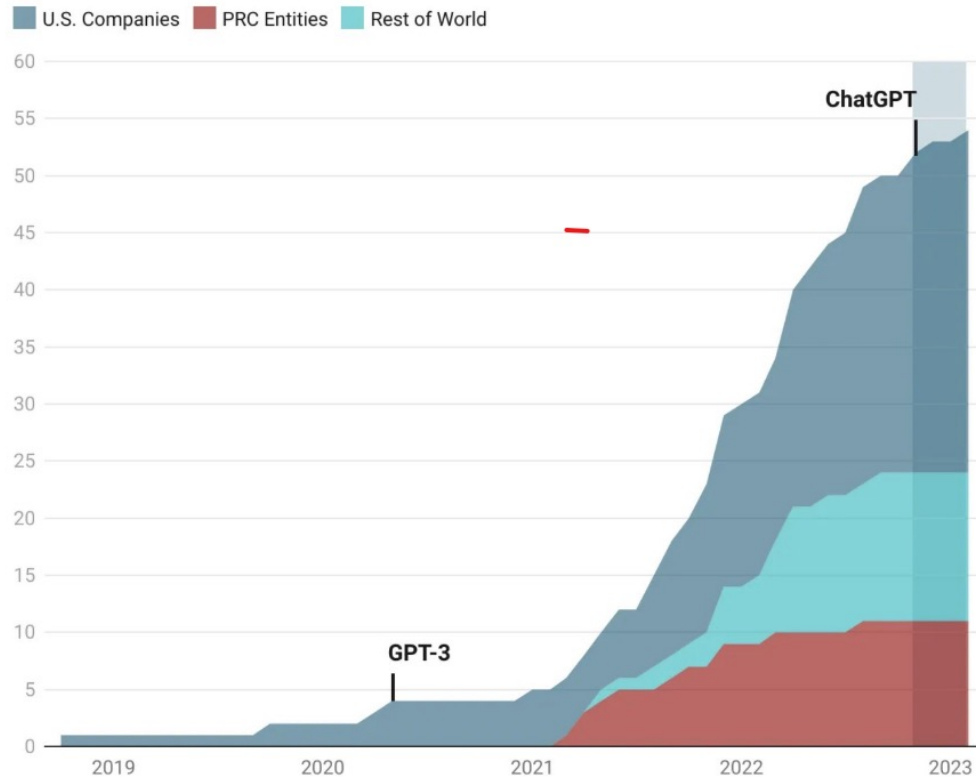
Source: McKinsey Global Institute analysis



Open data in a geopolitical context

An Accelerating Large Language Model Ecosystem

Total number of models ≥ 10 B parameters.



This dataset draws on SCSP research as well as data collected by Our World in Data. It includes new and updated LLMs with at least 10 billion parameters. It does not include tools or apps built by third parties on top of LLMs.

Chart: SCSP • Source: SCSP • Created with Datawrapper

China Delays Indefinitely the Release of G.D.P. and Other Economic Statistics

The unusual move comes as the country's ruling elite have gathered in Beijing for the twice-a-decade national congress of the Communist Party.

New York Times, 2022

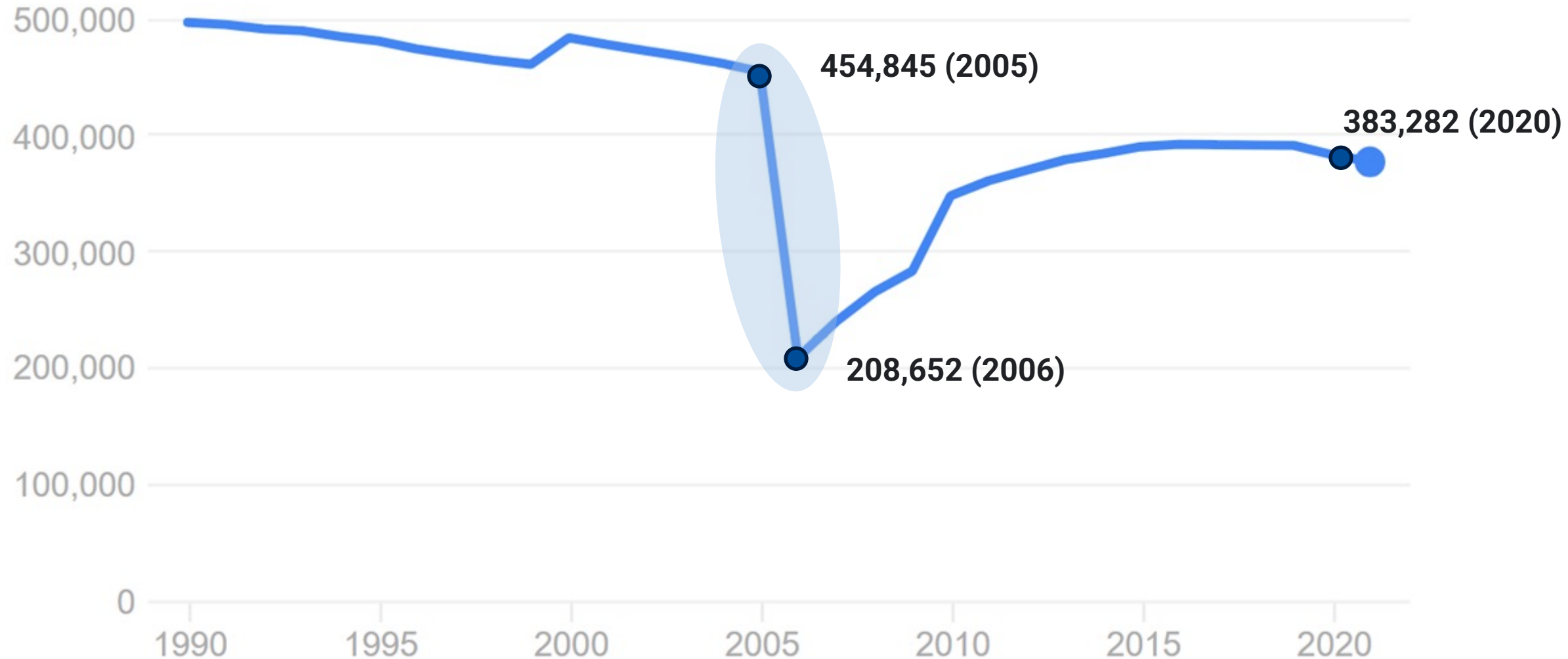








Population of New Orleans





**Through data, we see ourselves as part of
something much bigger**