

Data Democratization in the Age of AI



Oliver Wise | Chief Data Officer | US Department of Commerce



Open Data and AI: Putting it in Context



The conversations on GenAI in government

How can we promote responsible AI in society?

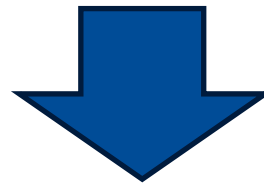
How do we leverage Gen AI responsibly within government?

How does government leverage AI to democratize access to data?



History is likely to repeat itself

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



The public will expect a natural language interface to accomplish many (most) tasks, especially data analysis

Credit: David Eaves



Photocopier -> FOIA Laws





API -> Open Data EO & Data.gov





Open data in the pandemic

JOHNS HOPKINS UNIVERSITY & MEDICINE | CORONAVIRUS RESOURCE CENTER

Home Topics ▾ By Region Events & News About Search Data by Region...

Tracking Home Data Visualizations ▾ Global Map U.S. Map Data in Motion Tracking FAQ

JHU Ceased Updates at: 3/10/2023, 8:21 AM
See Terms of Use for more info

Total Cases 676,609,951	Total Deaths 6,881,951	Total Vaccine Doses Administered 13,338,833,191
28-Day Cases 4,035,254	28-Day Deaths 28,018	28-Day Vaccine Doses Administered 28,156,730

Cases | Deaths by Country/Region/Sovereignty

US 28-Day: 959,794 9,451 Totals: 103,804,263 1,123,836	Japan 28-Day: 418,671 2,804 Totals: 33,329,551 73,046	Germany 28-Day: 355,168 2,275 Totals: 38,249,060 168,935	Russia 28-Day: 350,549 989 Totals: 22,086,064 388,521	Korea, South 28-Day: 290,039 396 Totals: 30,615,522 34,093	Taiwan* 28-Day: 216,931 778 Totals: 9,970,937 17,672	Brazil 28-Day: 170,852 1,613 Totals: 37,085,675 699,310	Austria 28-Day: 148,431 197 Totals: 5,961,143 21,970
---	--	---	--	---	---	--	---

Weekly Cases

Weekly Deaths

Weekly Doses Administered

Admin0 Admin1 Admin2 **28-Day** Totals Incidence Case-Fatality Ratio Global Vaccinations US Vaccinations Terms of Use

Weekly 28-Day

7

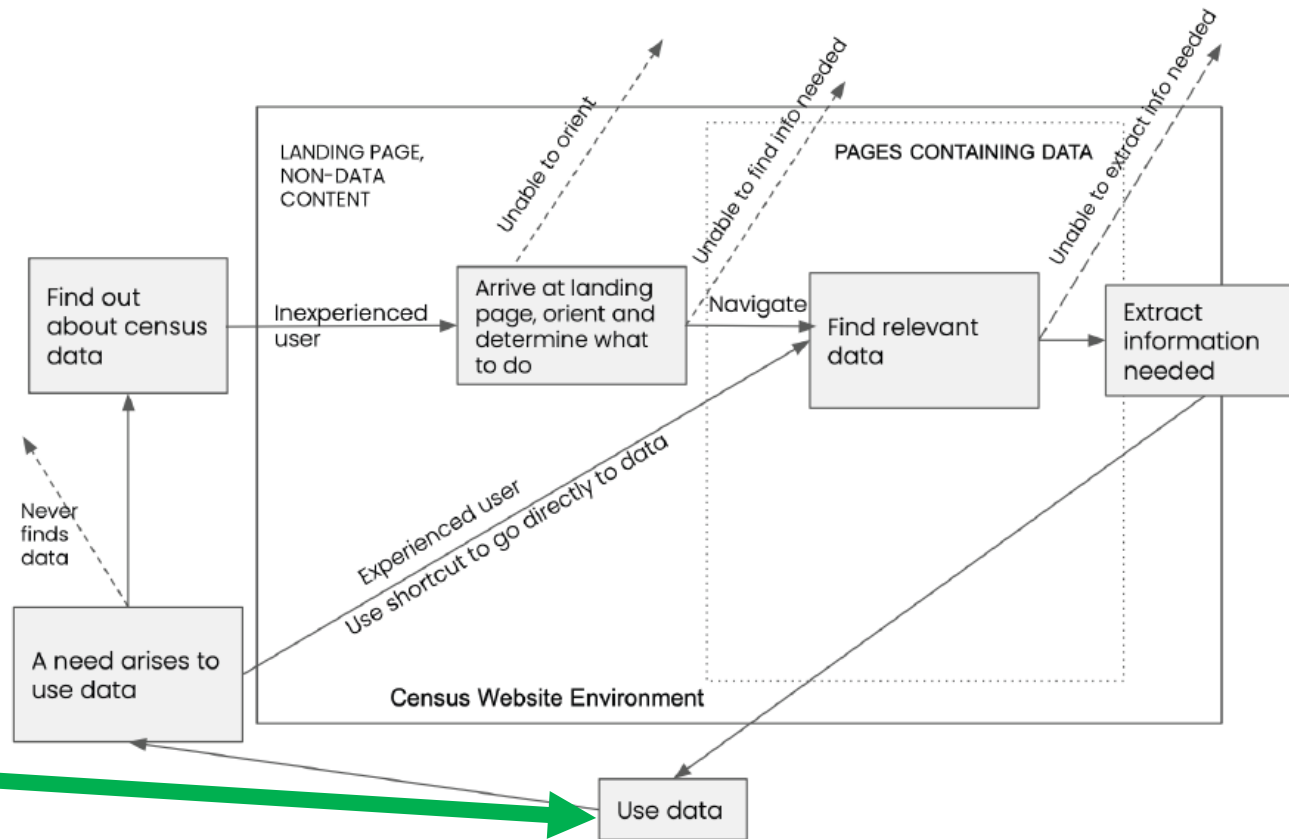


Open Data and AI



From this....

Figure 1 . General process for interacting with Census data



To this: user asks a question of chatbot

Credit: Drew Zachary, "Census data users and workflows: a cognitive task approach to understanding user their experience," 2023



However, GenAI hasn't gotten 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.

wrong²

wrong

wrong

true

wrong

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

true

wrong

true

wrong



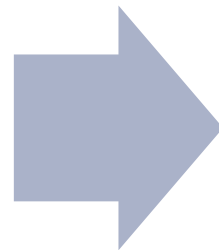
**We are not AI-Ready until our
data is.**



AI-readiness requires a shift in how government thinks data publishing standards

Machine readable

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

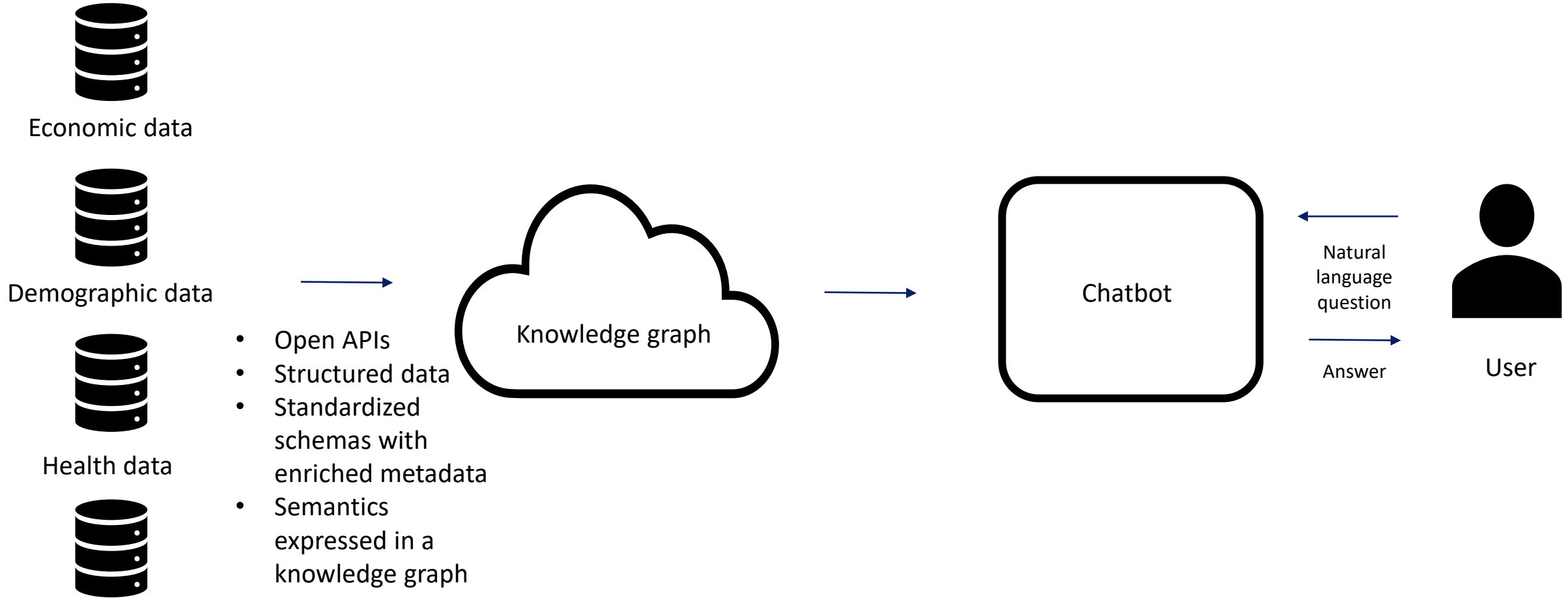


Machine understandable

- Structured data
- Standardized schemas with enriched metadata
- Semantics expressed in a knowledge graph



How do we make this work?



Credit: Swiss Federal Statistical Office



Problems to solve to realize this vision

- Policy guidance
- Capacity
- Fitness for purpose



So what?



Open data is foundational to our democracy

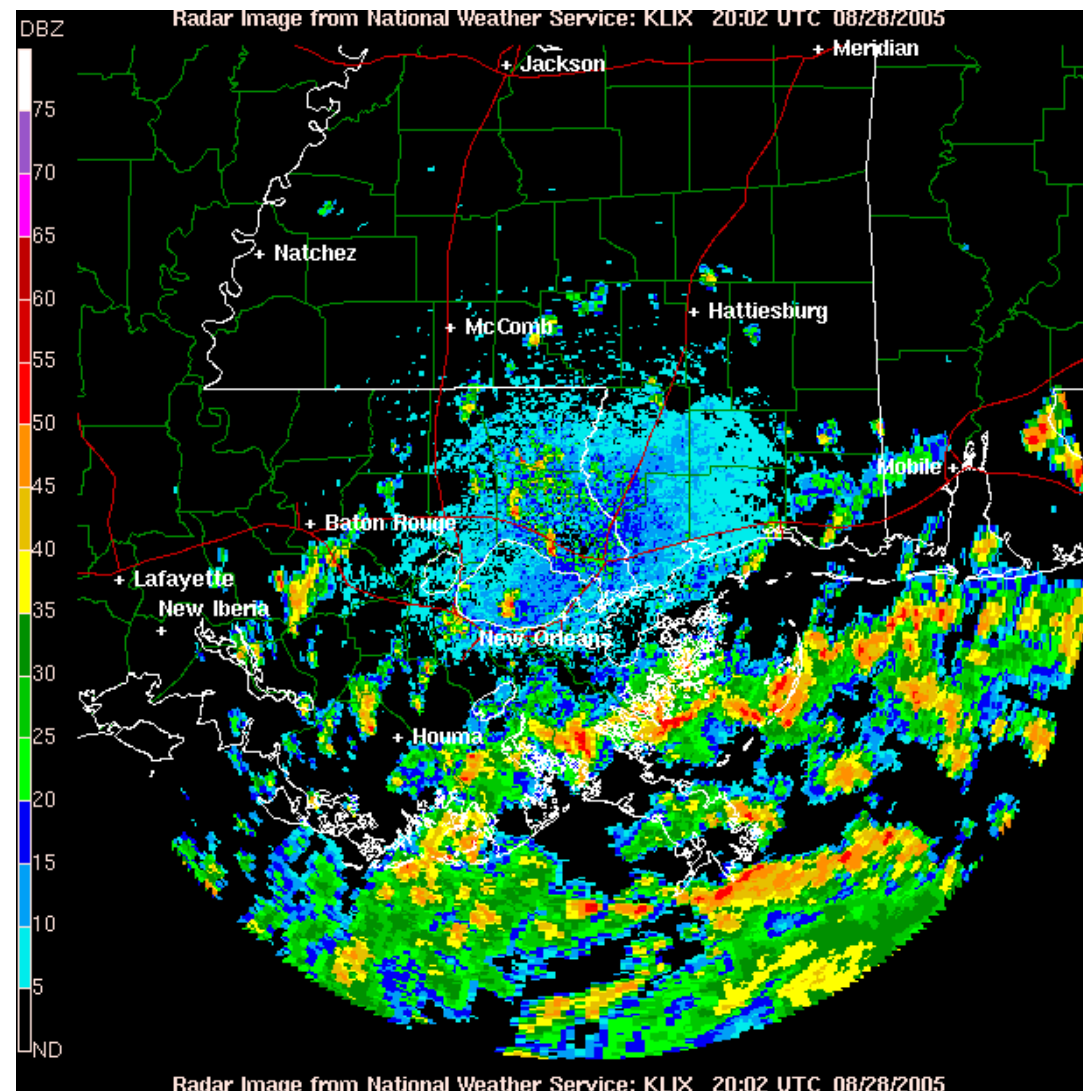




Open data drives scientific discovery

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.



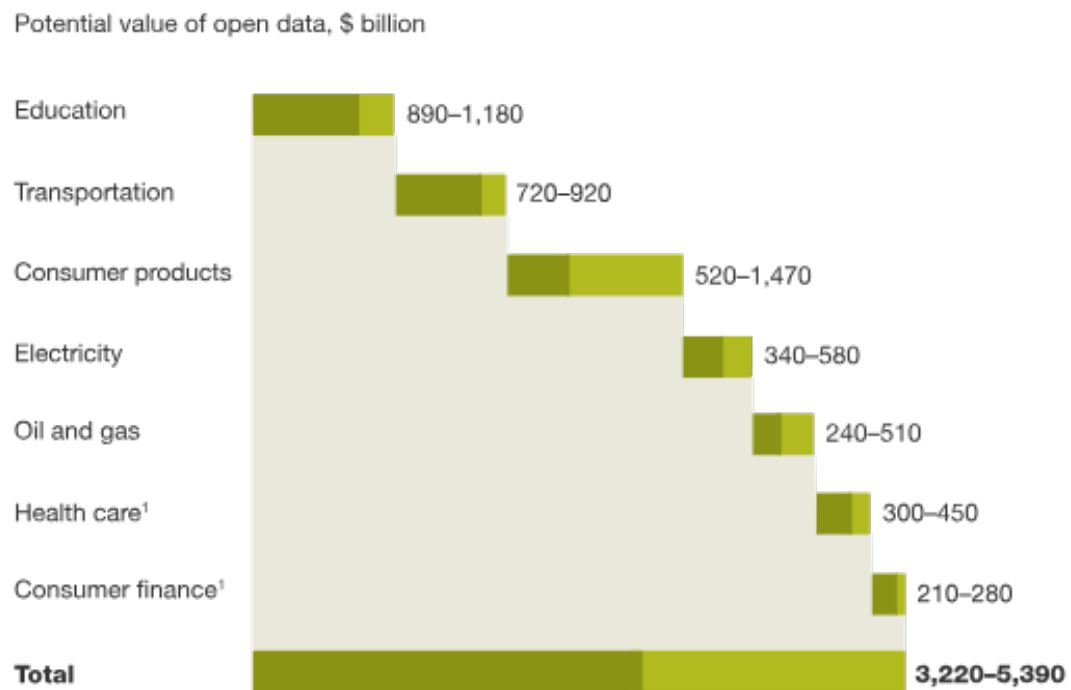
Radar Image from National Weather Service: KLIX 20:02 UTC 08/28/2005

National Weather Service Radar during Katrina (NOAA)



Open data is key to driving economic growth

Open Data can help unlock **\$3 trillion to \$5 trillion in economic value annually** across several sectors (McKinsey Global Institute, 2021)



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.

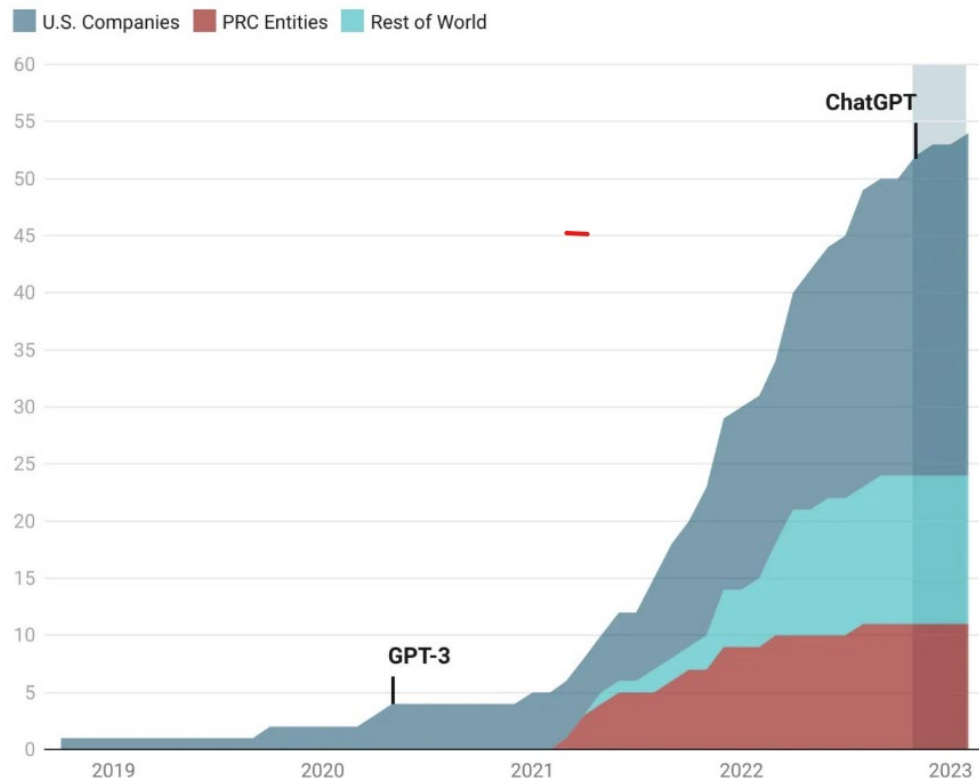
Source: McKinsey Global Institute analysis



Open data in the geopolitical context

An Accelerating Large Language Model Ecosystem

Total number of models $\geq 10B$ 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

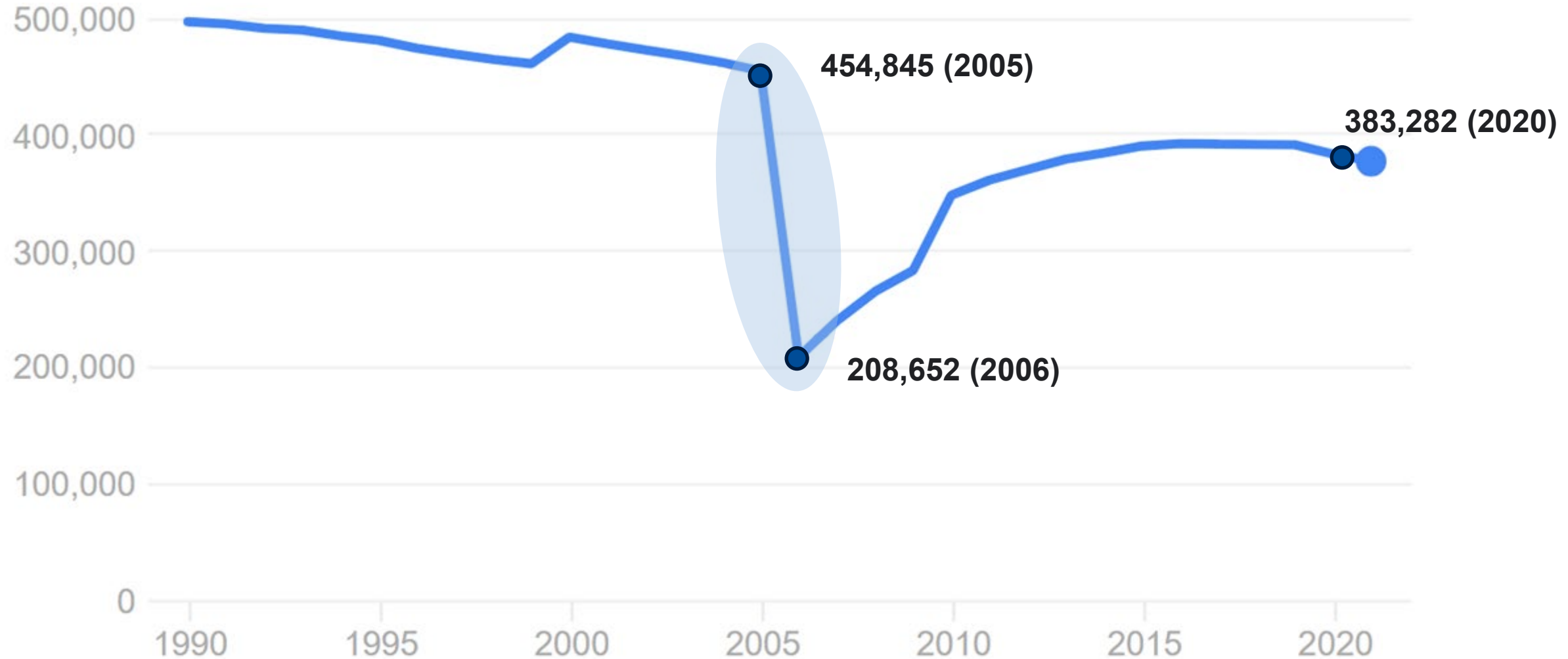








Population of New Orleans





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