

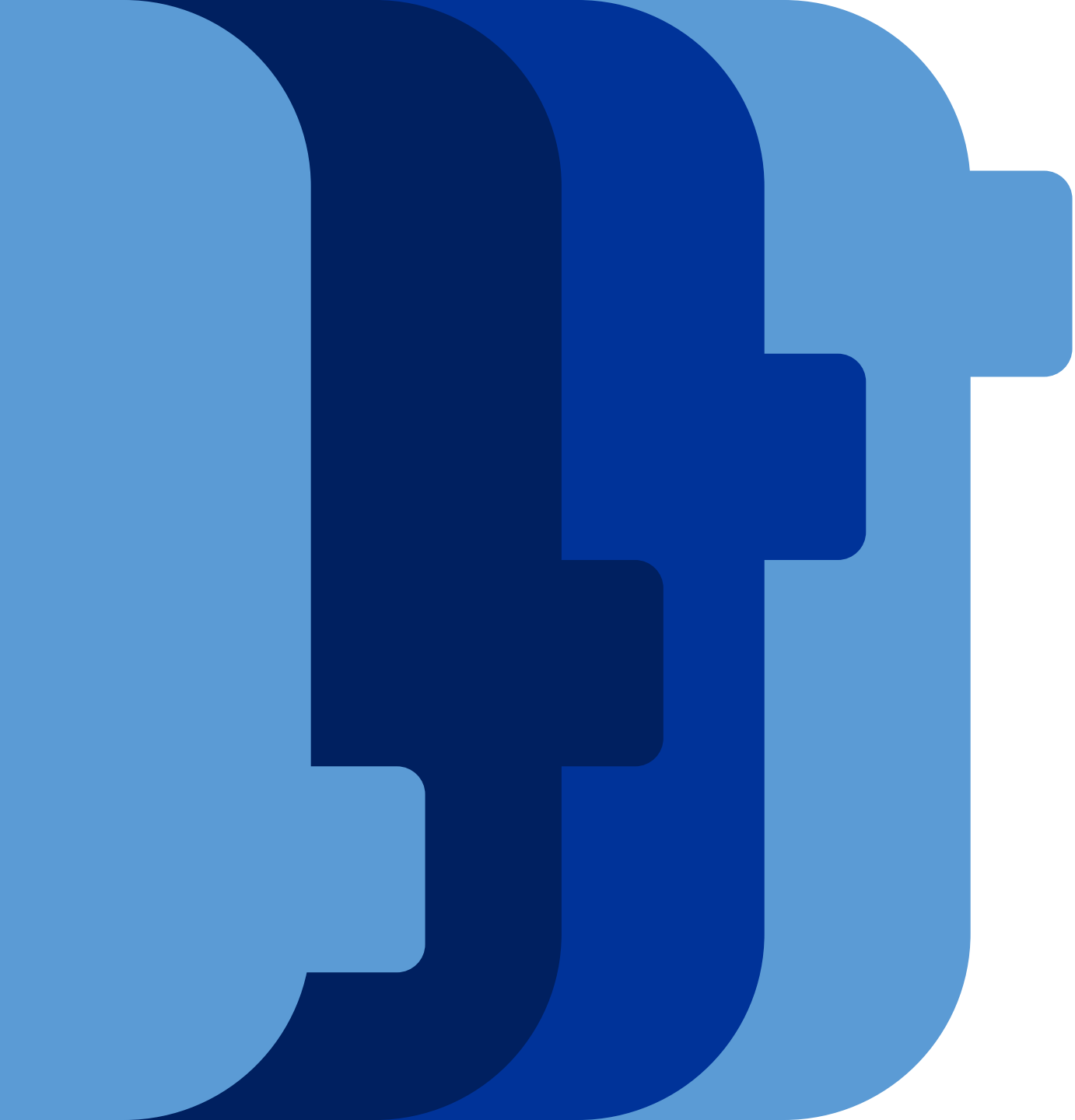
Data Science at the U.S. Census Bureau

Bowie State University Data Science & Analytics Symposium

Dr. Ron Jarmin

Deputy Director, U.S. Census Bureau

May 3, 2023





WHAT WE DO...

More than 130
Censuses and
Surveys...



Shape
your future
START HERE >

United States[®]
Census
2020



American Community Survey

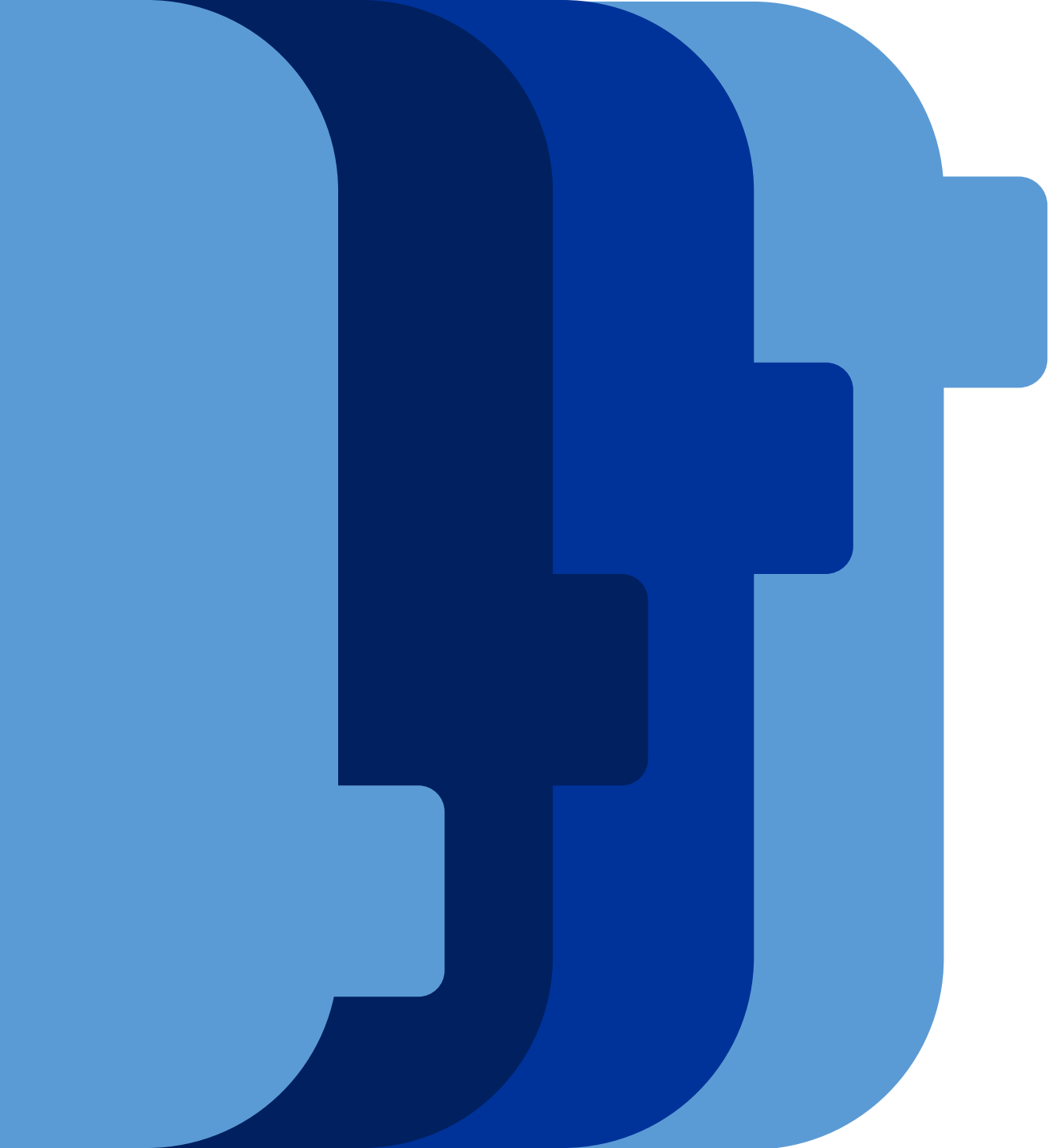
United States®
Census
Bureau

The **2022 Economic Census** is **Underway!**

[CENSUS.GOV/ECON](https://census.gov/econ)

United States[®]
Census
Bureau

The 2022
Economic
Census



AND...

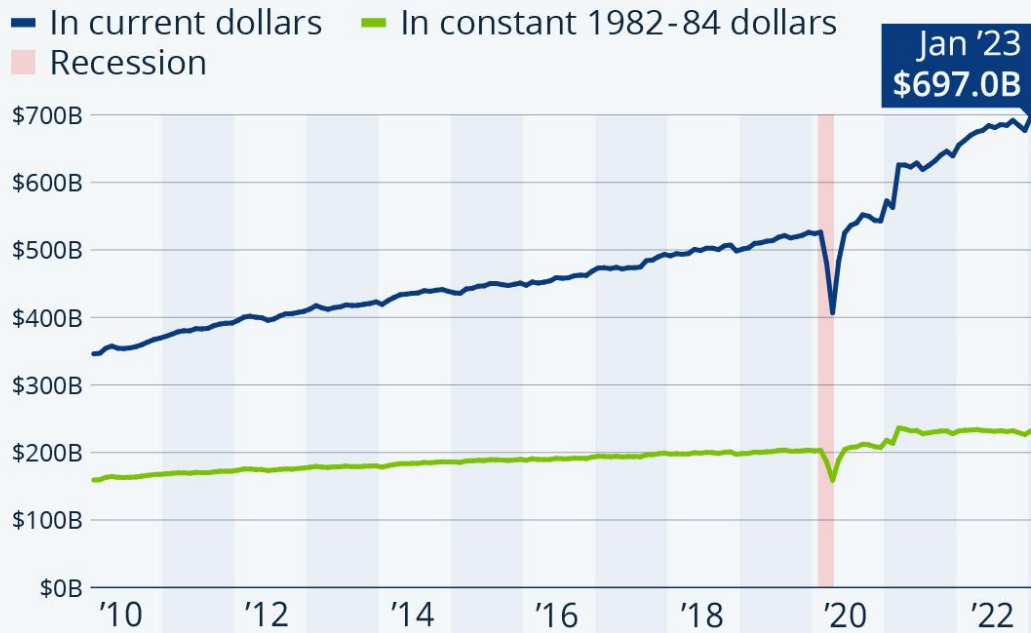


Official Poverty Measure

Poverty rate went up between 2019 and 2020.

Retail Sales Climb as Buyers Get Less Bang for Their Buck

Monthly retail and food services sales in the United States (seasonally adjusted)



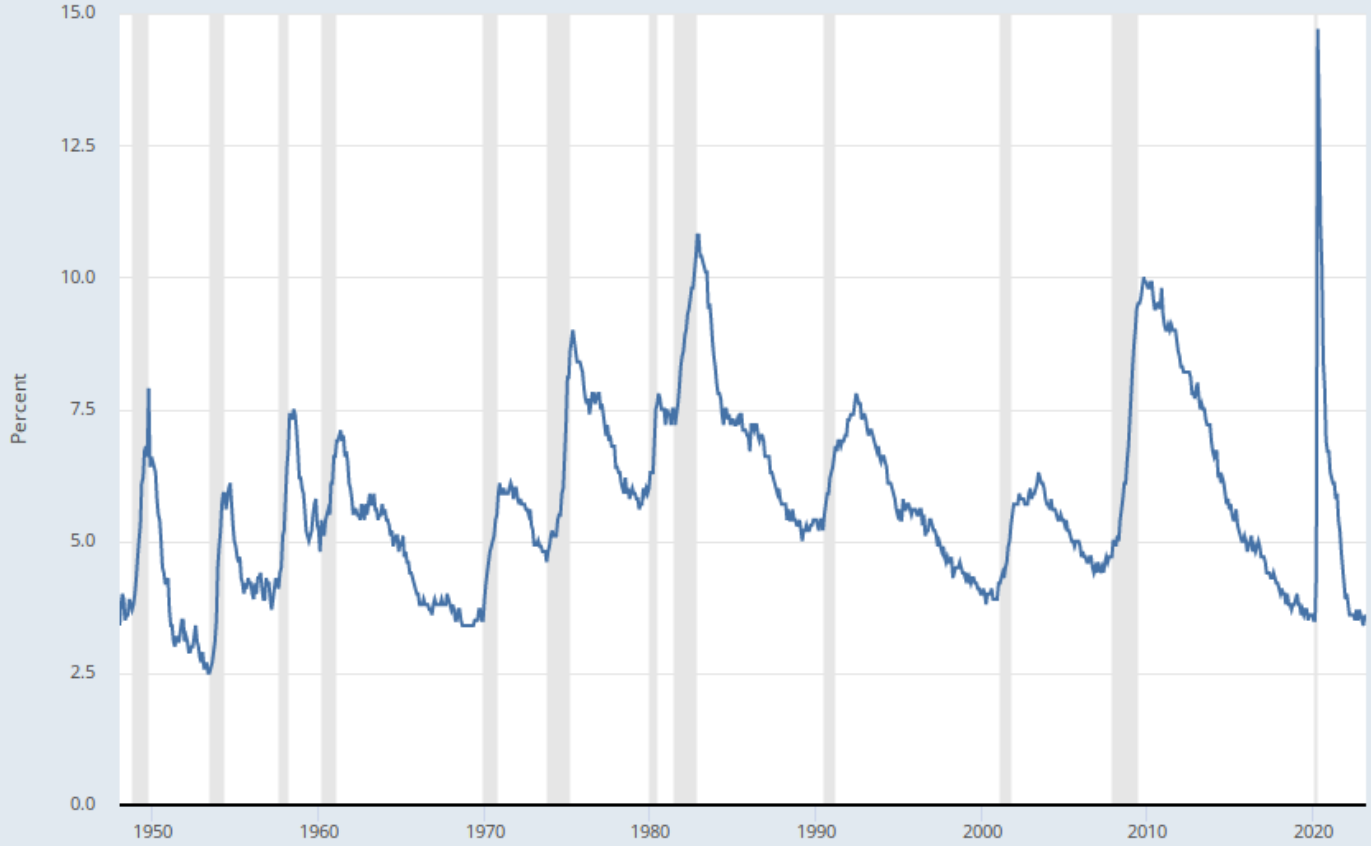
Source: U.S. Census Bureau



statista



FRED — Unemployment Rate

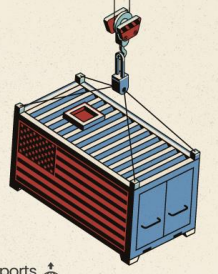


Source: U.S. Bureau of Labor Statistics

myf.red/g/12mls

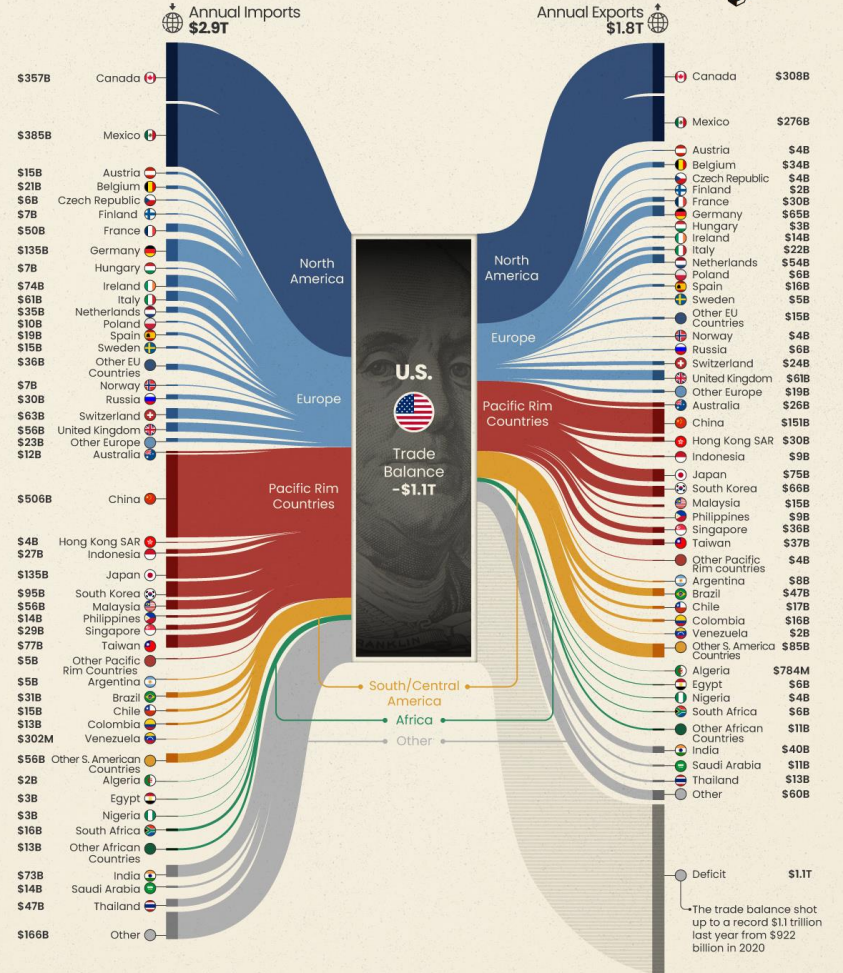


America's Trading Partners



Trade in Goods by Selected Countries and Areas in 2021

The U.S. economy grew 5.7% in 2021, the strongest since 1984. Goods exports hit an all-time high of \$1.8 trillion. But import growth dwarfed exports, shooting up the goods trade deficit to \$1.1 trillion.



The trade balance shot up to a record \$1.1 trillion last year from \$922 billion in 2020

Note: Country grouping by The U.S. Census Bureau



COLLABORATORS RESEARCH + WRITING Raul Amoros ART DIRECTION + DESIGN Sam Parker

SOURCE Census.gov

A hand holding a pen pointing at a tablet displaying various data points and icons. The background is a light blue grid with various icons and numbers. The icons include a globe, a laptop, a document, a padlock, and a mail icon. The numbers are 207.70, 210.95, 207.7, 210, 149, and 23.26. The text "Modernizing Federal Statistics" is overlaid in white.

Modernizing Federal Statistics

Data Ecosystem for Official Statistics

20th Century

- Data is expensive
- Digitization is scarce but growing
- Dominant role of federal statistical agencies for economic measurement
- Government surveys contribute most primary source data

21st Century

- Data is cheap(er)
- Digitization is ubiquitous
- Increasing role of the private sector in economic measurement
- Administrative and third-party data increasingly important

Challenges to the existing 20th Measurement Paradigm

- Declining survey/census response rates – increasing costs.
- Increased demand for statistics – more timely, more granular, and in expanding domains – difficult to cost effectively meet demand with surveys alone.
- The proliferation of alternative third-party source data.
- Lack of resiliency for key data collection platforms during large shocks.
 - Pandemic, natural disasters, social unrest impact survey operations (2020 Census, ACS, CPS).
 - Large economic shocks can impact performance of survey questions (e.g., CPS).
 - Shocks can delay delivery of administrative data.

Improvements in Economic Measurement Increasingly Based on Administrative and/or Third-Party Data

- Federal examples
 - QCEW (BLS), LBD/BDS/BFS (Census)
 - QWI and OnTheMap (Census)
- Private Sector examples
 - ADP Jobs report
 - JPMorgan Chase Institute.
 - Census Monthly State Retail Sales

But, also from New(ish) Surveys

- JOLTS (BLS)
- Census Pulse Surveys

Blending Data for 21st Century Economic Measurement

- Goal: Use alternative third-party data sources, coupled with survey and administrative data, to produce more timely, more detailed, and less burdensome products.
- As in...
 - New monthly, state-level retail sales estimates
 - Achieved by blending survey data, administrative tax data, and data obtained through a third-party data aggregator
 - 2020 ACS
 - Non-response bias addressed by modeling with admin data

How do we build the 21st Century Economic Measurement System?

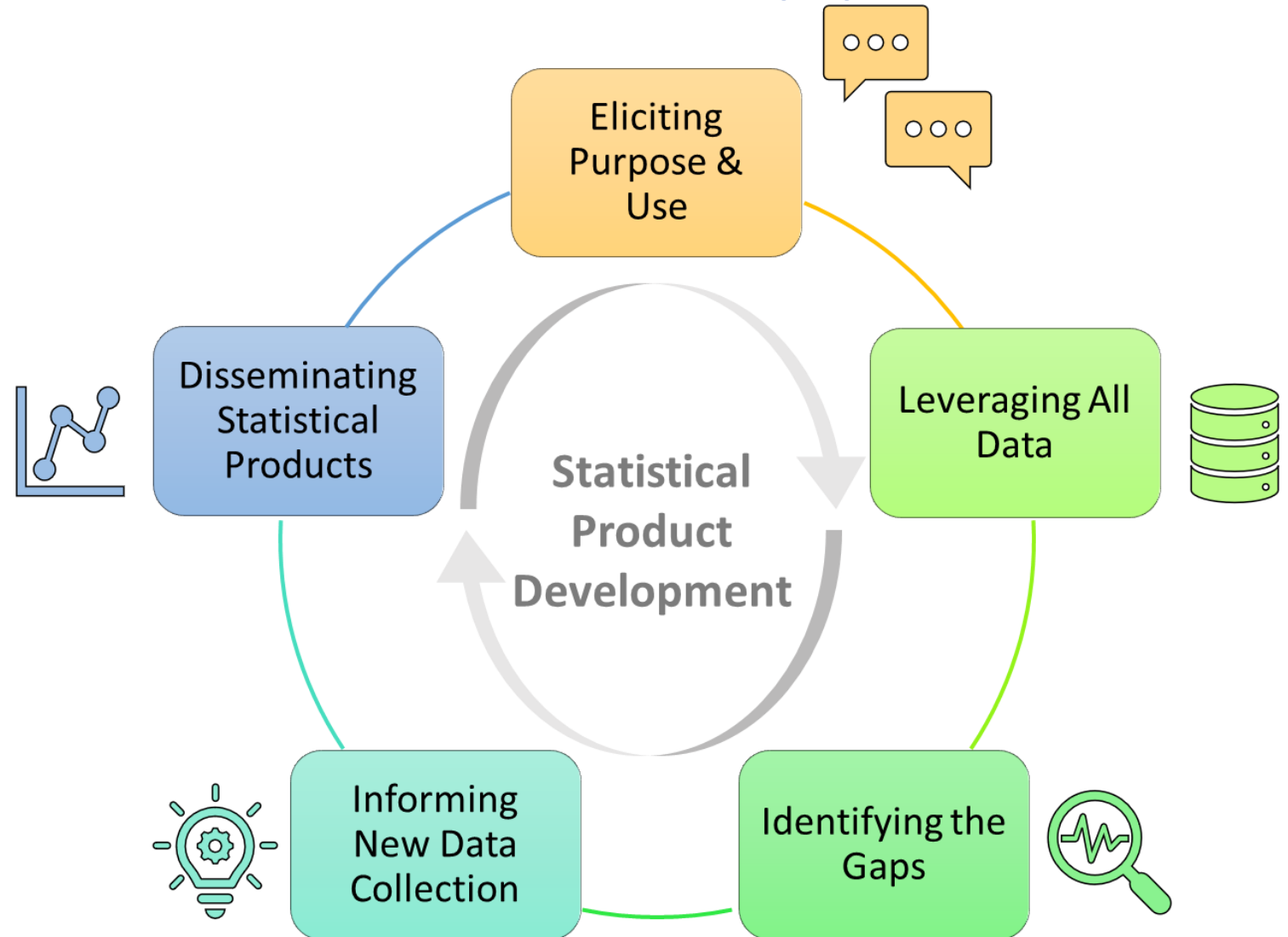
- Source Data
 - Role of businesses and households (survey respondents vs data sharing)
 - Role of aggregators (monetization?)
 - Role of statistical agencies
- Methods and Computation
 - Value Chain of statistical production – role of engineers
 - How do ensure transparency and reliability?
- Public Provision of Statistics
 - Private vs Public vs Academic

Statistical Product First Approach

Flipping the Focus

Determine what information stakeholders need to reach their objectives

From there, shape the statistical products to be developed



Key Components of this Approach



DATA VISUALIZATION AND STORYTELLING

*The graphical
representation of data
into usable insights*



INDUSTRY DOMAIN

*Having expertise in a
specific domain which
is helpful in navigating
the data*



USER EXPERIENCE/ USER INTERFACE

*The process of
enhancing user
satisfaction
with a product*



ADVANCED MATH AND STATISTICS

*Applying mathematical
theory and statistical
methods to analysis*

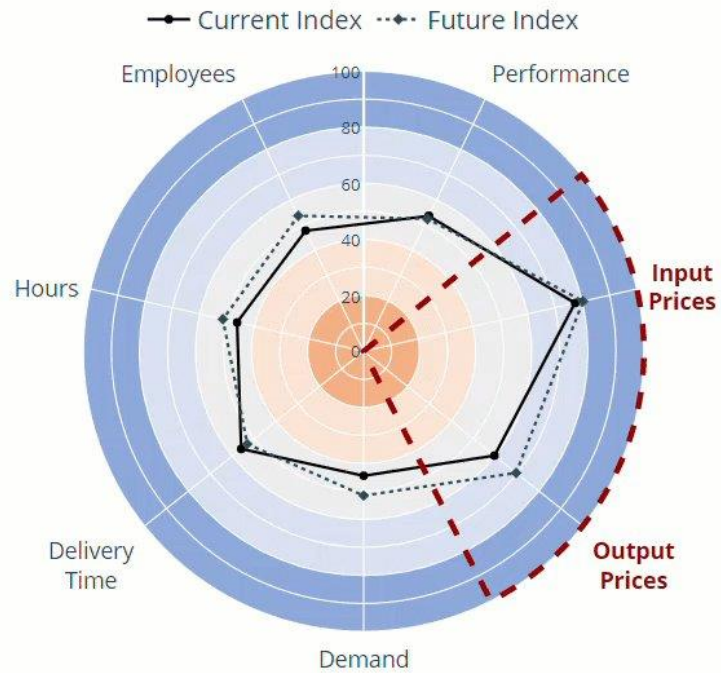


DATA ENGINEERING AND PROGRAMMING

*Learning methods to
enhance data-centric
processes*

Examples:

Business Trends and Outlook Survey – Paired Indexes



National Average

Index = 0: Most Widespread Decreases, Index = 100: Most Widespread Increases

Data Collected: 10/24 - 11/6

occur, recovery depends on the community's ability to withstand the effects of the event. In order to facilitate disaster preparedness, the Census Bureau has developed new small area estimates, identifying communities where resources and information may effectively mitigate the impact of disasters.

Variation in individual and household characteristics are determining factors in the differential impact of a disaster. Some groups are less likely to have the capacity and resources to overcome the obstacles presented during a hazardous event. Resilience estimates can aid stakeholders and public health officials in modeling these differential impacts and developing plans to reduce a disaster's potential effects.

Individual and household characteristics from the 2019 American Community Survey (ACS) were modeled, in combination with data from the Population Estimates Program to create the CRE.

Risk factors from the 2019 ACS include:

- Income to Poverty Ratio
- Single or Zero Caregiver Household
- Crowding
- Communication Barrier
- Households without Full-time, Year-round Employment
- Disability
- No Health Insurance
- Age 65+
- No Vehicle Access
- No Broadband Internet Access

Initial Release Date

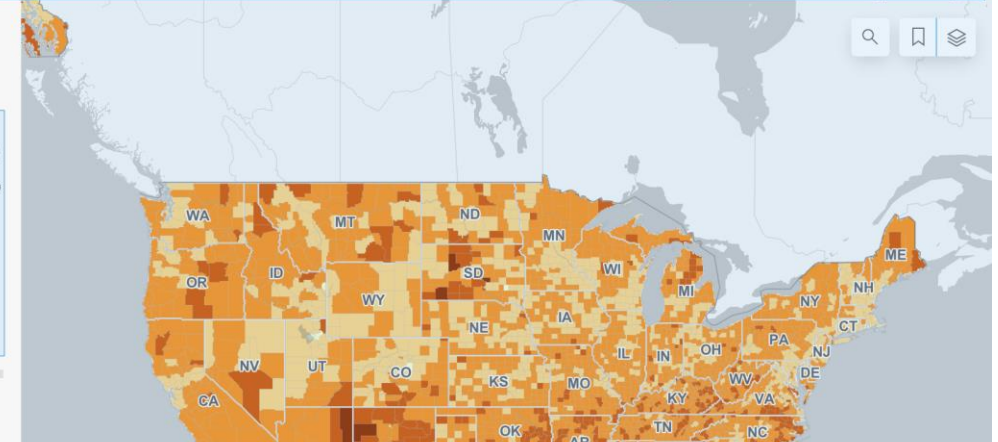
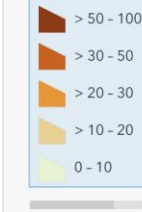


Thematic Risk Factor (RF) based on the risk factor selected above

Use Layer List below to turn on and off supplemental layers

Thematic Risk Factor (Counties) 2019

Estimated Population (%)



PSEO

What is PSEO? Download Data Email Us Tutorial Help

Data Type
 Earnings Flows

State
 Virginia

Institution
 Norfolk State University

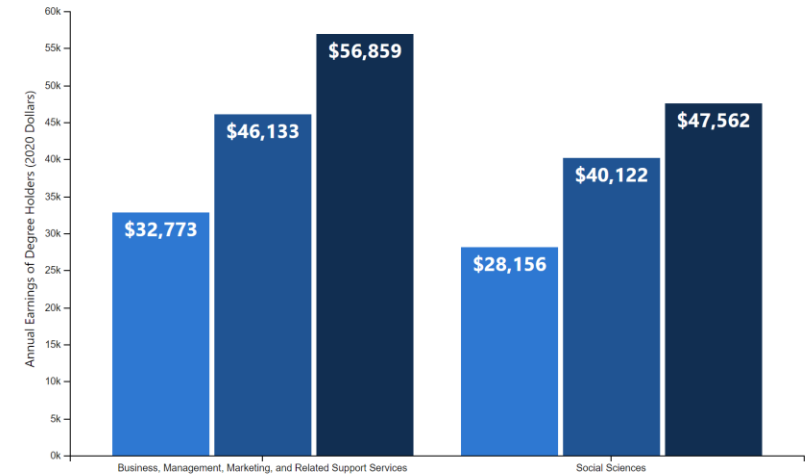
Degree Level
 Baccalaureate

Graduation Cohort
 All Cohorts

Compare Earnings by
 Time Percent

Earnings-percentile Filter
 25 50 75

Legend
 1 Years Postgrad
 5 Years Postgrad
 10 Years Postgrad



Program Specificity
 General Detailed

Sort Program List by
 Size Name

All Instructional Programs	Business, Management, Marketing, and R...	Multi/Interdisciplinary Studies
Health Professions and Related Programs	Social Sciences	Psychology
Communication, Journalism, and Related ...	Public Administration and Social Service P...	Visual and Performing Arts

Transformation and Modernization

- Needs modern tools and infrastructure
- Needs a diverse and passionate team with modern skills

Tools and Infrastructure – Unique to the Census Bureau

Creating a Business Ecosystem to Modernize the Census Bureau's Statistical Foundation

Data Input
Categories

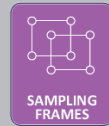


Data
Ingest
and
Collection
for the
Enterprise
(DICE)

Processing
and
Manage-
ment
of
Incoming
Data

Enterprise Data Processing, Computing
and Management (Enterprise Data Lake)

Other datasets, databases, sampling frames, etc.
(derived from and/or linkable to Enterprise Frames)



Enterprise Frames
(i.e., foundational datasets)



Internal
Products

Enterprise
Data
Dissemi-
nation
(CEDSCI)



External
Products

Data User
Feedback
and New
Questions
and Needs
for Data

Identify
New
Data, New
Sources,
New
Products

Technical and Subject Matter Skills Needed

- Data Science
- Computer Science
- Software Engineering
- Cybersecurity
- Statistics
- Economics
- Demography
- Geography / GIS

In Fed Speak....

- **Data Scientist** – 1560: Degree or combination of education and experience in mathematics, statistics, computer science, data science or field directly related to the position.
- **Mathematical Statistician** – 1529: Degree or combination of education and experience, including at least 24 semester hours in mathematics and statistics (at least 12 hours in mathematics and 6 in statistics).
- **Survey Statistician** – 1530: Degree or combination of education and experience, including at least 15 semester hours in mathematics and statistics (at least 6 hours in statistics).
- **Economist** – 0110: Degree in economics with at least 21 semester hours in economics and 3 semester hours in statistics, accounting, or calculus.
- **Information Technology (IT) Specialist** – 2210: Degree in computer science, engineering, information science, or related field, including at least 24 semester hours in the IT field.
- **Computer Scientist** – 1550: Degree in computer science or another program another with 30 semester hours in a combination of mathematics, statistics, and computer science courses.
- **Geographer** – 0150: Degree in geography or related physical or social science fields, including at least 24 semester hours in geography or related fields.

Thank You