FINAL PROGRAM

Government Advances in Statistical Programming (GASP)
Tuesday, 24 June – Friday, 27 June 2025

Click here to JOIN the conference.

Tuesday, 24 June 2025

12:00 pm Conference Opening Session

Welcome by Co-Chairs: Lisa M. Frehill (Department of Energy, Retired)
Peter B. Meyer (Bureau of Labor Statistics)

Reflections on Government Open-Source Software: A Conversation with Dr. Joseph Castle

Participants:
Joseph Castle (Former Director, Code.gov, U.S. General Services Administration)
José Bayoán Santiago Calderón (Bureau of Economic Analysis)

Attendees should put their questions in the Zoom Q&A feature.

12:50 – 1:25 Lightning 1: Dealing with Nonresponse and Missing Values

Chair: Peter B. Meyer (Bureau of Labor Statistics)

Developing a Python Package for Multiple Imputation by Chained Equations (MICE)
Angelica Phillips, Joey Marshall, **Besufekad Alemu**, Ruth E. Sarafin (All from U.S. Census Bureau)

Sensitive Data, Smarter Solutions: The Role of Instrument Test Paradata
Renee Ellis and Ana Sanchez Rivera (Both from U.S. Census Bureau)

Recursive Computation via an Information Sieve
Tucker McElroy and Redouane Betrouni (Both from U.S. Census Bureau)

Item Response Theory as a Model-Based Alternative for Assessing Item Nonresponse
T. Ryan Johnson (U.S. Census Bureau)

1:25 – 2:45 Paper Session 1: Adapting to the Real World

Chair: Benjamin Schneider (Westat)

Automating Multilingual Census Data Processing: A Transformer-Based Pipeline for Efficient Language Detection and Translation for Short-Text
Hannah Zimmerman and Renee Ellis (Both from U.S. Census Bureau)

Use of Farm Machine Data to Inform Establishment Statistics at USDA NASS
Sean Rhodes, Dave Johnson, Michael Gerling, Chad Garber, Darcy Miller, Denise Abreu (All from USDA - National Agricultural Statistics Service)

Move Slowly and Break Things: A New Paradigm for National Forest Inventory
Andrew Lister (USDA Forest Service, Northern Research Station); John Hogland (USDA Forest Service, Rocky Mountain Research Station)

Predicting 'Yes': ML & Diverse Data to Boost Respondent Cooperation
Rashi Saluja, Ryan Hubbard, Jill Carle, Hanyu Sun, Gizem Korkmaz, and Brad Edwards (All from Westat)

Open Census: A Federal Statistical Agency's Approach to Open Science
Emily Molfino (U.S. Census Bureau)

2:45 – 3:15 Lightning 2: Analyzing Data Quality

Chair: Laura Cutrer (NOAA, retired)

Using Machine Learning to Identify Key Predictors for Invasive Mold Surveillance
Zainab Salah (Centers for Disease Control (CDC)), Samantha L Williams (CDC), Brendan R Jackson (CDC), Sebastian Wurster (University of Texas), Jose A Serpa (Baylor College of Medicine), Carolyn Z Grimes (University of Texas), Robert L Atmar (Baylor College of Medicine), Tom M Chiller (CDC), Dimitrios P. Kontoyiannis (University of Texas), Luis Ostrosky-Zeichner (McGovern Medical School), and Mitsuru Toda (CDC)

Using Simulation Techniques to Evaluate Variance in the Low Response Score
Ian Le and Ralph Culver III (Both from U.S. Census Bureau)

Using Health Data and Text Mining to Improve Cost Recovery: The SUMAR Program
Ariana Bardauil, Eugenia Villanueva, **Dacio Martínez**, Federico Di Tata (All from Buenos Aires City Government), Manuel Rodríguez Tablado (Hospital Italiano)

3:15 – 3:30 Announcements and BREAK

3:30 – 5:00 Workshop 1: Navigating Tough Conversations in Statistical Collaboration

Presented by:
Dr. Emily Griffith Professor of the Practice and Associate Head, Department of Statistics, NCSU
Dr. Julia Sharp, Acting Statistical Engineering Division Chief, NIST
Organizer: Dr. Donna LaLonde, Associated Executive Director, American Statistical Association

Statistical practitioners face difficult conversations in their interactions with their clients and collaborators. This workshop will build participants' confidence to effectively communicate with clients and customers when challenging topics or situations arise.

5:00 – 6:00 Social Hour **Zoom Link will be shared during the announcements.**

Wednesday, 25 June 2025

12:00 – 12:05 pm Day 2 Welcome / Announcements

12:05 – 1:00 pm Paper Session 2: Small Area Estimation

Chair: Matt Williams (RTI)

Challenges in Small Area Estimation: Rare Characteristics and Variance Estimates
Danny Friel, Michael Hudak, Michael Lettau, Erin McNulty, Kelly Rossman, Richard Uhrig, **Xingyou Zhang**, Yu Zhang (All from Bureau of Labor Statistics)

Small Area Estimation Workflow in the Census Integrated Research Environment
Stas Kolenikov (NORC at the University of Chicago)

A Machine Learning Approach for Counting Language Minority Groups in the United States
Joseph Kang and **Adam Hall** (Both from U.S. Census Bureau)

1:00 – 1:35 pm Lightning Talks 3: Finding, Matching, and Classifying

Chair: Michael Stanley (Analytical Mechanics Associates/NASA)

Matching Legacy Time Series Models with Implied Marginals of Multivariate Models
Soumadeep Bhowmick (University of Maryland, Baltimore County (UMBC)), James Livsey (U.S. Census Bureau), Anindya Roy (UMBC/U.S. Census Bureau)

Bootstrapped Standard Errors for Monthly Estimates for Monthly Injuries Rates
Benjamin Raymond (Bureau of Labor Statistics)

Bayesian Person-Place Model: Probabilistic Address Imputation from Administrative Records
Nathan Welch (MITRE)

Split-Apply-Combine with Dynamic Grouping: R Package accumulate
Mark van der Loo (Statistics Netherlands and Leiden University)

1:35 – 1:45 BREAK

1:45 – 3:05 pm Paper Session 3: Protecting Privacy

Chair: Matt Williams (RTI)

Privacy-Preserving Model Auditing: Definitions, Techniques, and Tradeoffs
Tomo Lazovich and Michael Walton (Both from U.S. Census Bureau)

Privacy-Preserving Autocoders for Survey Response Classification
Robert Chew (RTI), Matthew R. Williams (RTI), Elan A. Segarra (Bureau of Labor Statistics, BLS), Alexander J. Preiss (RTI), **Amanda Konet** (RTI), David Oh (BLS), Erin Boon (BLS), Terrance D. Savitsky (BLS)

Introducing TaCo: A Python Tool for Assessing and Augmenting Suppression Methods
Elan Segarra (Bureau of Labor Statistics)

Noise Injection for Automated Disclosure Avoidance
Mikaela Meyer (MITRE), Saimun Habib (MITRE), Gary Benedetto (U.S. Census Bureau), Rolando Rodriguez (U.S. Census Bureau), Jordan Awan (Purdue University), Giuseppe Germinario (U.S. Census Bureau), Jordan Stanley (U.S. Census Bureau), Evan Totty (U.S. Census Bureau), Richmond Stevenson (MITRE)

Entropy as a Measure of Disclosure Risk for Tabular Data
John Grant, Luca Sartore (National Institute of Statistical Sciences), and Alex Tarter (All from USDA - National Agricultural Statistics Service, NASS)

3:05 – 3:50 pm Lightning Talks 4: Publishing Data and Metadata

Chair: Raza Lamb (U.S. Census Bureau)

Sub-Sampling as Data Protection: A Case Study of the Asian American Survey
Jennifer Taub and Ben Reist (Both from NORC at the University of Chicago)

ARcenso: A Package Born from Chaos, Powered by Community
Andrea Gomez Vargas (rOpenSci, INDEC Argentina) & **Emanuel Ciardullo** (R en Buenos Aires, INDEC Argentina)

A Data Model to Facilitate Research for Federal TANF Data
Molly Rossow and Emily Wiegand (Both of NORC at the University of Chicago)

Maximizing Linkage in Address Data: Spatial, Exact, and Fuzzy Matching
Timothy Champney (MITRE), Hongxun Qin (MITRE), and Stephanie Coffey (U.S. Census Bureau)

Parameter Estimation in Record Linkage Through Dimensional Reduction
Daniel Weinberg and Yves Thibaudeau (Both from U.S. Census Bureau)

3:50 – 5:00 pm	Paper Session 4: Data Editing and Imputation
Chair: Wendy L. Martinez (U.S. Census Bureau)	
Looking for Dirty Data in NYC Open Data **David Tussey** (New York City Department of Information Technology, retired) and Jun Yan (University of Connecticut) *A Semi-Supervised Active Learning Approach for Block-Status Classification*. **Atul Rawal,** James McCoy, Andrew Duvall, and Elvis Martinez (All from U.S. Census Bureau) *Editing Survey Data Using an R Shiny Dashboard* **Francisco Cifuentes Villarroel** (Energy Information Administration) *On the Use of Machine Learning Methods for Missing Data Problems* **Sixia Chen** and Chao Xu (Both from University of Oklahoma Health Sciences)	

5:00 – 5:05 pm Day 2 Announcements

Thursday, 26 June 2025

10:30 – 11:45 pm Workshop: Quarto - To Tell Your Story with Data

Presented by: **Isabella Velásquez** (Senior Product Marketing Manager, Posit PBC)
Organizer: **Gwynn Gebeyehu** (Internal Revenue Service)

Join us for a 75-minute introduction to Quarto, the next-generation scientific and technical publishing system. In this workshop, you'll learn the fundamentals of Quarto. We'll cover document creation, code integration, and output customization, equipping you to build your own Quarto documents.

https://bit.ly/gasp2025-quarto

12:00 – 12:05 pm Day 3 Welcome and Logistics Review

12:05 – 1:20 pm Paper Session 5: Blending Data

Chair: Clayton Knappenberger (U.S. Census Bureau)

A Discussion of Privacy Preserving Record Linkage Methods
Emily Gentles (RTI)

A New Python Package for Fast, Open-Source Geocoding and Record Linkage
Alex Lee (University of Melbourne)

Enhancing Data Visualization and Accessibility on data.census.gov
Faith Whittington and **Tyson Weister** (Both from U.S. Census Bureau)

Blended Data to Measure Income for Tribal Areas
Randy Akee (UCLA, U.S. Census Bureau), Cass Dorius, Carla Medalia, Sallie Keller (All from U.S. Census Bureau)

Enhancing the Current Population Survey with Administrative Tax Data
Max Ghenis and Nikhil Woodruff (Both from PolicyEngine)

1:20 – 1:55 pm Lightning Talks 5: Natural Language Processing

Chair: Raza Lamb (U.S. Census Bureau)

gtrendshealth: An R Package to Access a Google Trends Service for Public Health Investigation
Oscar de León (Centers for Disease Control)

Algorithm Development to Estimate Pregnancy Data from Electronic Health Records
Carolina Mengoni Goñalons (Buenos Aires City Ministry of Health, University of Buenos Aires), Juliana Reves Szemere (Buenos Aires City Ministry of Health, National University of San Martin, National Pedagogical University), María Cristina Nanton (Buenos Aires City Ministry of Health, University of Buenos Aires)

Table Manners for RAG: Dining on Machine-Understandable Official Statistics
Michael Long, Timothy Navarro, Alexander J. Preiss, Lauren Klein Warren (All from RTI)

Applying Semantic Search for Census Working Papers
Eric Zou, Ethan Crouse, **Rahul Ramakrishnan** (All from Virginia Tech)

Comparing Web-Scraped Establishment Survey Frames of Industrial Hemp Growers
Michael Gerling (NASS), **Chad Garber** (NASS), Tyler Wilson (NASS) and Katherine Vande Pol (Smithfield Foods)

1:55 – 2:05 pm Break

2:05 – 3:20 pm Paper Session 6: Large Language Models (LLMs) at Work

Chair: Clayton Knappenberger (U.S. Census Bureau)

Total Recall? Evaluating the Macroeconomic Knowledge of Large Language Models
Leland Crane (Federal Reserve), **Akhil Karra** (Carnegie Mellon University), Paul Soto (Federal Reserve)

Synthetic LLM-Generated Texts to Train Small Models for Automatic Coding
Adrián Pérez Bote (Spanish Statistical Office), Sebastián Gallego Herrera (Spanish Statistical Office), Andrés Jurado Prieto (Spanish Statistical Office), Carlos Sáez Calvo (Spanish Statistical Office)

LLM Assisted Causal Knowledge Graph Generation Framework for Survey Data
Atul Rawal (U.S. Census Bureau) & Richard Martinez

Leveraging Survey Metadata for LLM Reasoning via Knowledge Graphs
Irina Belyaeva (U.S Census Bureau), Chris Carino (U.S Census Bureau), Liang-Chi Wang (U.S Census Bureau)

3:20 – 3:50 pm Lightning Talks 6: Classification and Sensing

Chair: Rachel Sloan (U.S. Census Bureau)

Statistical Classifications: A FAIRy Tale
Daniel Gillman and **Peter B. Meyer** (Both from Bureau of Labor Statistics)

From Canopy Cover to Context: Combining Traditional Forest Inventory and Remote Sensing to Better Represent Trees in Urban Areas
Alexander Young (University of New Hampshire)

Evaluating MLLM's for Noise Reduction in Cloudy-Region Geospatial Analysis
Sagnik Chakravarty (University of Maryland)

3:50 – 4:55 Paper Session 7 (Contributed): Modernizing Efforts for Survey Data Analysis

Organizer and Chair: Matt Williams (RTI)

Session Overview: Survey samples underpin many official statistics programs. In this session we present about integrating current best practices into modern data analysis and dissemination pipelines. For example, we describe the creation of a small area estimation program for health surveys and connecting a traditional survey analysis software (SUDAAN) to Python pipelines. We will also present recent and upcoming approaches to design and analysis of data collections and modelling to increase efficiencies over the traditional collection and analysis of survey data. This includes working with multi-level models for variance decomposition for survey data and augmenting probability samples with intentionally designed non-probability samples.

Creating a Small Area Estimation and Dissemination Pipeline
Matt Williams (RTI) and **Marcia Underwood** (RTI)

Integrating SUDAAN into Modern Analysis Pipelines
Victoria Scott (RTI) and David Wilson (RTI)

Challenges with Multilevel Modelling and Uncertainty Estimation for Survey Data
Hunter McGuire (RTI) and Matt Williams (RTI)

A Framework for Combining Probability and Non-probability Samples to Over-sample Rare Populations
Taylor Lewis (RTI) and Mahmoud Elkasabi (RTI)

FRIDAY, 27 June 2025

12:00 – 12:05 pm Day 4 Welcome and Logistics Review

12:05 – 12:50 pm Lightning Talks 7: Shiny for Sharing **Chair: Gwynn Gebeyehu (Internal Revenue Service)**
Connecting Experts to the Data: Using R Shiny to Improve Child Welfare in TN **Alexis Fleming** and Rameela Raman (Both from Vanderbilt University) *R Shiny for the Visualization of Large Multi-Source Datasets and Data Integration: Use Case of Spanish Tourism Statistics* **Jordi Verdú Naranjo** (Statistics Spain), Diego De la Puente Alonso (University of Valladolid) *Economic Indicator Analysis Tool* **Axel Kent**, Rachel Butler, Alex Prevatte, Tandin Dorji, Alisha Gurnani, Eric Muro, Benjamin Griffis (All from U.S. Census Bureau) *A Shiny App for Linked Micromaps* **Randall Powers** (Bureau of Labor Statistics) and Wendy Martinez (U.S. Census Bureau) *Enhancing Ecce Signum Framework Usability with a Graphical Interface* **Redouane Betrouni** and Tucker McElroy (Both from U.S. Census Bureau)

12:50 – 1:50 Closing Session

Panel Discussion **Converting Legacy Code to R and Python – Experiences and Hints and Tips from the Census Data Analytics and Research Group**
Moderator: Renee Ellis (U.S. Census Bureau) **Panelists**: **Ana Sanchez Rivera** (U.S. Census Bureau) **Hannah Zimmerman** (U.S. Census Bureau) **Kelyvette Ortiz Fontanez** (U.S. Census Bureau)
As part of overall data modernization efforts, many agencies are converting code used for a variety of projects to open-source software. Join us for a discussion of the process of converting code from the perspectives of a small research group at Census. We will share our experience converting existing legacy code from a variety of projects, challenges of code conversions, and some hints and tips.

1:50-2:00 Closing Remarks

Ellen Galantucci (Veterans Affairs) and Chair, Data Science for Federal Statistics FCSM Interest Group