GASP

Government Advances in

Statistical Programming 2018

October 24th, 8:00am-4:30pm

October 25th, 8:00-1:00pm

Janet Norwood Conference and

Training Center

U.S. Bureau of Labor Statistics

2 Massachusetts Avenue, NE

Washington, DC 20212

Conference and Conference Center Information

Janet Norwood Conference Center, Ground Floor

U.S. Bureau of Labor Statistics

2 Massachusetts Ave, NE

Washington, DC

20212

To RSVP please visit on or before October 22nd: https://www.eventbrite.com/e/gasp-government-advances-in-statistical-programming-tickets-50961320682

All visitors check in with the security desk and are required to show a government-issued photo ID (i.e., driver's license, passport). All visitors must pass through a metal detector. We recommend you schedule time to account for security screening.

Once screened, proceed to the elevator bank and take the elevator down to the Ground Floor. The Janet Norwood Conference Center is around the corner to the left and through the glass doors. Please check in at the registration desk.

Special thanks to our organizers and sponsors: Interagency R Users Group, the Federal Committee on Statistical Methodology (FCSM) Computational Statistics Interest Group, and the Washington Statistical Society.

Day 1

Early Morning Sessions: 9:00 am to 10:30 am

Theme: From Research to Reality—Rooms 7 and 8

How do new ideas get put into place? Who creates, operates, and maintains the software and programs at our respective agencies? Do R programs belong in production? In this session, we look at the pipeline from research to production and address questions that may shape our choices of software.

- On Implementing a New Imputation Method into Production in the 2017 Economic Census Illustrated through Selected Vignettes
  Jenny Thompson – US Census Bureau

- Variance Estimation for Product Sales in the 2017 Economic Census: Challenges in Implementing Multiple Imputation-Based Variance Estimation
  Matthew Thompson – US Census Bureau

- Updating a Generalized Imputation System to Include a Quadratic Program
  Laura Bechtel – US Census Bureau

- Using R in the Production of Official Statistics at NASS
  Darcy Miller – NASS USDA

Theme: Take Steps Toward Reproducibility—Rooms 9 and 10

The talks in this session are themed around making workflows more transparent and reproducible. Speakers discuss the use of R, RStudio, and other open source tools to make that happen.

- Continuous Integration and Deployment for R
  Zech Kottilil – Westat

- Using "git" to leverage development and external collaboration
  William Sexton – *US Census Bureau*

- Organizing research projects with an efficient open-source tool (emacs org-mode)
  Feiming Chen – FDA

- Open Source Software for Automating and Ensuring Quality in Official Statistics: An Example Using R and RStudio for Stratified Simple Random Sampling
  Darryl Creel – RTI

Late Morning Sessions: 11:00 am to 12:30 pm

Theme: Programming for a Purpose—Rooms 7 and 8

In this session we examine scripts, functions, and systems incorporating R to address a variety of specific research questions or agency needs.

- Using Gaussian copulas to generate a synthetic population
  Yijun Wei — NISS and USDA/NASS

- Indicator for Top-Coding Effects on a Household Survey Income Elasticity of Demand Estimates
  Danny Yang – BLS

- Using R for Advanced Statistical Modeling in the National Energy Modeling System
  Janice Lent – EIA

- Transparent and Reproducible Research in Agricultural Official Statistics
  Andreea Erciulescu – NISS and USDA/NASS

Theme: Dashboards and Interactive Tools—Rooms 9 and 10

The presenters discuss prototyping, building, and deploying dashboards and interactive tools using R Shiny and other related software packages. Some of the tools discussed may be public facing, others are being considered for internal use.

- R shiny platforms, Statistical Web Applications from Booz Allen Hamilton
  Zach Blumenfeld – Booz Allen Hamilton

- Polishing Analysis Tools Until They Are Shiny
  Andrew Dau — USDA/NASS

- Spatial Analysis with R and Power BI: Mapping U.S Census Data
  Tony McGovern – em data

- Designing, building, and deploying an interactive survey monitoring dashboard using R Shiny
  Joe Murphy – RTI

Lunch: 12:30 to 1:30 pm (on your own)

Plenary Session: 1:30 pm to 2:30 pm

Theme: Briefing from the Federal Data Strategy Development Team

Afternoon Sessions: 3:00 pm to 4:30 pm

Theme: Lightning Round!—Rooms 7 and 8

This session contains a group of short talks and poster walkthroughs on a variety of topics. Come be part of the discussion!

- Automated Document Generation using R Markdown
  Eric Budge – Insight Policy Research

- R Applications for Non-Data Reporting Offices
  Suzanna Stephens –Independent

- Cleaning and Analyzing Data from the U.S House of Representatives
  M Daniel Turse – Independent

- Analyses of Chicago Subway (L) Ridership
  M Daniel Turse – Independent

- Attendance and Graduation Rates in High Schools Across the US
  Helen Levy-Myers – Athenas Workshop

- Using NLP, the Census' Commodity Flow Survey team
  Christian Leonard Moscardi – US Census

Theme: Pulling It Together!—Rooms 9 and 10

R packages that put federal data in the public's hands, methods for linking a variety of data sources, and environments that could change the way we survey. Come see how data gets pulled together.

- There's a package for that? R Packages that Support Analyses of Common Federal Datasets
  Kelsey Gray – Insight Policy Research

- Linking Public Data Sources to Create Localized Official Statistics
  Dipak Subedi – USDA/ERS

- Estimating Traffic Crash Counts Using Crowdsourced Data - Pilot analysis of 2017 Waze and police accident reports in Maryland
  Dan Flynn – DOT

- Encrypted processing on SGX enclave hardware in a commercial cloud
  Cavan Capps – US Census Bureau

Day 2

Morning: 8:30 am to 10:00 am

Theme: Applications of Machine Learning—Rooms 7 and 8

This session examines applications of machine learning, natural language processing, computer vision and the software environments that support them.

- Using Computer Vision to Process Vehicle Dashboard Displays in Transportation Safety Research
  Kristin Chen – Westat

- Exploring Author Influence with Networks, NLP, and Trend Analysis
  Benjamin Ulloa – Independent

- Semiparametric deep learning from crop yield modeling, using Keras/Tensorflow
  Andrew Crane-Droesch – USDA/ERS

- Improving Estimates of Arrest Related Deaths with Machine Learning
  Peter Baumgartner – RTI

Theme: R from Multiple Perspectives—Rooms 9 and 10

Speakers share their perspectives and experiences on the use of R. Compare and contrast the uses of R in the research setting, in regulatory environments, and in a federal statistical agency.

- MISSUITE: A Shiny Application for Clinical Trial Missing Data Analysis
  Chenguang Wang—Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University

- Using R in a regulatory environment: some FDA perspectives
  Paul Schuette- FDA

- INCA: an R package for Integer Calibration
  Luca Sartore- NISS and USDA/NASS

Late morning: 10:15 am to 11:45 am

Theme: Bayesian Analysis with R—Rooms 7 and 8

In this session we examine a variety of Bayesian methods in the areas of time series, spatio-temporal analysis, small area estimation, and survey analysis. We explore the use of R and related software to facilitate those computation.

- An Imputation Model for the Advanced Monthly Retail Trade Survey using Hierarchical Bayesian Regression Models in STAN
  Stephen Kaputa – US Census Bureau

- Modeling Spatio-Temporal Trends using R
  Ali Arab – Georgetown University

- Using R for Bayesian Analyses of Survey Data
  Matthew Williams – National Science Foundation

- Fitting a Bayesian Fay Herriot Model
  Nathan Cruze – USDA/NASS

Theme: Demonstrations—Rooms 9 and 10

Our speakers demonstrate a variety of capabilities that have been prototyped and developed using open source tools.

- Accounting for All Federal Surveys
  Jared Teeter — Independent

- Shiny applications without Shiny
  Emily Mitchell – AHRQ

- Shiny and Geospatial maps
  Brandon Kopp – BLS

- Shiny Demonstration Utilizing Bootstrap Approach to the Application of First Digits Analysis
  Matt Corrigan – BLS

Plenary Session: 12:00 am-1:00 pm

Marck Vaisman, Data Community DC

Announcements and Adjourn