

# GASP 2019

# Government Advances in Statistical Programming workshop

September 23<sup>rd</sup>, 2019, 8:00am-4:30pm

Janet Norwood Conference and Training Center  
U.S. Bureau of Labor Statistics  
2 Massachusetts Avenue, NE  
Washington, DC 20212

**Thanks to our supporters and sponsors:** Washington Statistical Society,  
Interagency R Users Group, and the  
Federal Committee on Statistical Methodology (FCSM) Computational Statistics Interest Group



## Introductions

8:00 - 8:30am Check in

8:30 - 8:45am Introductions, Learning to Network *Mike Jadoo (BLS)*  
*Robert Brandau (Career Confidence)*

8:45 - 9:00am Check-in, networking game to meet two new people

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

Session 1A: Using R to streamline analysis Rooms 1 and 2

- Who is more likely to submit a 311 request, *Monica Puerto (American University)*
- Generic Drug Evaluation and R-package SABE, *Elena Rantou (US Food and Drug Administration)*
- Using Amazon Web Services in R with the Paws package, *David Kretch (Summit Consulting)*
- Using the Current Population Survey to Answer DOL-Relevant Questions: Applications in R, *Samuel Rowe (Department of Labor's Office of the Assistant Secretary for Policy)*

Session 1B: Model selection/validation with statistical software Rooms 7 and 8

- A glimpse into the "Electricity Initiative" at the U.S. Energy Information Administration, *Greg Lawson (Energy Information Administration)*
- Tableau for Data Scientists, *Joel Hutchison (Tableau)*
- Using Cross-Validation for Variable Selection in Generalized Linear Models with Complex Survey Data, *Darryl Creel (RTI International)*
- Scalable Approximate Kernelized Logistic Regression in R with an Application to the Census of Agriculture, *Jonathan Abernethy (USDA-NASS)*

**Break** 10:30 to 10:45 am

Network with a presenter - ask about their project, meet a new person

**Late Morning Sessions:** 10:45 am to 12:15 pm

Session 2A: Python libraries for text analysis and to streamline production Rooms 1-2

- Easy deep learning text classification with gobbli, *Jason Nance (RTI International)*
- Using Natural Language Processing and Machine Learning to Process Open Text Field Comments in a Panel Study, *Kristin Chen (Westat)*
- (SLIDES NOT POSTED) Using Active Learning Framework and Automated Machine Learning Algorithms to Code Open-ended Responses, *Rosalynn Yang (Westat)*

Session 2B: Using R and open source to automate for efficiency      Rooms 7 and 8

- R Data Packages of Multi-Reader Multi-Case Studies and Simulation Tools to Support the Development of Reader Performance Evaluation Methods, *Si Wen (FDA)*
- Hierarchical models in the production of official statistics: a discussion of some practical aspects, *Andreea Erciulescu (Westat)*
- Topic Modeling Consumer Complaints for Risk Analysis, *Benjamin Bloom (Federal Reserve Board)*
- Beyond the Analysis: An Approach to Operationalizing Your Results using Open Source JavaScript Frameworks, *Dylan Holt (Accenture Federal)*

Session 2C: Lightning Round! Room 3

- Complex Survey Variance Estimation and Design Effects in R using the Rstan and Survey packages, *Matt Williams (NCSES/NSF)*
- Responses of SRISK to Changes in Bank Capital Regulation, *Elizabeth Duncan and Benjamin Kay (Federal Reserve Board)* ~~WITHDRAWN~~
- Ensemble Models with Basketball, *John Thomas (George Washington University)*
- Automating the Maintenance of Survey Frames, *Elizabeth Willhide and Keith Finlay (Census Bureau)*
- Text Analysis of Death Certificate Records to Ascertain Drugs Involved in Deaths from the National Vital Statistics System, *Merianne Spencer and Brigham Bastian (NCHS)*

**Lunch:** 12:15 to 1:15pm    Lunch, on your own.

The BLS cafeteria is on the first floor, and Union Station has many restaurants.

**Early afternoon sessions:** 1:15 to 2:45 pm

**Session 3A: Data dashboards for quick monitoring and analysis** Rooms 1 and 2

- SMART: An Open Source Tool to Facilitate Auto-Coding, *Caroline Kery (RTI International)*
- (SLIDES NOT POSTED) Power Up NASS Data Analytics with MS Power BI, *Leanne Tang (USDA-NASS)*
- Think Outside of the Box(plot), *Jerry Valerio (Tableau)*
- Modern Techniques for Exploring Text Data, *Peter Baumgartner (RTI International)*

**Session 3B: Open source software for fast information** Rooms 7 and 8

- Faster Computation for Hierarchical Bayesian-Model Using Rcpp Packages, *Lu Chen (USDA-NASS)*
- Asking Consumers about their Finances, *Kimberly Kreiss and Mike Zabek (Federal Reserve Board)*
- Using Python, PostgreSQL and R to analyze NIBRS data from the Crime Data Explorer, *Ian Thomas (RTI International)*
- Weight Calibration across Packages, *Dr. Stanislav Kolenikov-(Abt Associates)*

**Session 3C: Lightning Round!** Room 3

- Create Interactive Motion Charts Using R Package GoogleVis, *Bidong Liu (Data and Analytic Solutions, Inc.) and Zhengyi Fang (Social & Scientific Systems, Inc.)*
- Automating Occupational Profiles Using RMarkdown, *McLeod Brown (BLS)*
- Using R to Automate the Development of Custom Reports in Word Format, *Alexandra Gates and Mina Zheng (NORC at the University of Chicago)*
- Linking EEOC Case Investigation Records by Employer Name using Text Analytics, *Ada Harris (U.S. Equal Employment Opportunity Commission)*
- Rapid Implementation of Test Design Using Python, *David Oh (BLS)*

**Break** 2:45 to 3:00pm

Network with a presenter - ask about their project, meet a new person

## Late afternoon sessions, 3:00 to 4:15pm

### Session 4A: Application of Python and other software for production – Rooms 1 and 2

- Analyze, visualize, and . . . itemize: Tax policy analysis with Tax-Cruncher, *Peter Metz (American Enterprise Institute)*
- Synergy between remote sensing and machine learning for crop yield prediction, *Luca Sartore (USDA-NASS)*
- Getting Started with Using Amazon Mechanical Turk to Label Data, *Brandon Kopp (BLS)*

### Session 4B: Analysis using R-shiny and other open source software      Rooms 7 and 8

- Using R and Shiny Applications to Analyze and Visualize NAEP Assessment Data, *Emmanuel Sikali, Michael Lee, Paul Bailey and Ting Zhang (American Institutes for Research and NCES)*
- A Tour of USDA NASS's Decision Support System, *Nathan Cruze (USDA-NASS)*
- Measuring the cost and impact of open source software as intangible capital, *Gizem Korkmaz (University of Virginia)*

### Session 4C: Lightning Round!

Room 3

- Adaptivity of Media and Behavioral Modeling, *Mitchell Shuey (Independent)*
- (SLIDES NOT POSTED) What is the meaning of this? How natural language processing can analyze unstructured data, *Kelsey Gray (Insight Policy Research)*
- Routine Automation and Replication of PowerPoint Presentations in Python, *Elaine Wilcox-Cook (Insight Policy Research)*
- An Exploratory Research on Optimization of CFS Sampling Design, *Mehdi Hashemipour (BTS)*
- Science of Visual Analysis – Beyond Statistics, *Jerry Valerio (Tableau)*

## Announcements and Adjournment, 4:15 - 4:30pm

Rooms 1 and 2

## Remote Access Information by Session

**AUDIO:** Audio information can be found once you access the meeting. Note that you can dial-in for audio only, if necessary.

- Audio on the WebEx sessions (2C, 3C, 4C: Lightning Rounds) is all via the telephone with the same number and access code used for all. The number is given here: **1-866-865-9536, 744 124 3 #**
- Audio on the Zoom sessions (all of the rest) can be via telephone or the computer. **PLEASE** use the computer for audio, if possible. If you use the phone for Zoom sessions, then the phone number for all sessions is **1-929-205-6099** or **1-669-900-6833**. Only the access codes change for each session.
- If remote access isn't working for you we may be able to help on the spot; email [meyer.peter@bls.gov](mailto:meyer.peter@bls.gov), [Jadoo.michael@bls.gov](mailto:Jadoo.michael@bls.gov), and [Martinez.wendy@bls.gov](mailto:Martinez.wendy@bls.gov).
- Please send slides by Friday Sept 20. You may bring an update on a flash drive but it adds difficulty.

### **9:00 AM to 10:30 AM**

Session 1A     [Using R to streamline analysis](#) 666126207 #  
Session 1B     [Model selection/validation with statistical software](#) 722857413 #

### **10:45 AM to 12:15 PM**

Session 2A     [Python libraries for text analysis](#) 977147246 #  
Session 2B     [Using R to automate for efficiency](#) 373757563 #  
Session 2C     [Lightning round](#) **Audio:** 1-866-865-9536, 744 124 3 #

### **1:15 PM to 2:45 PM**

Session 3A     [Data dashboards](#) 437101314 #  
Session 3B     [Open source software](#) 384613050 #  
Session 3C     [Lightning round](#) **Audio:** 1-866-865-9536, 744 124 3 #

### **3:00 PM to 4:15 PM**

Session 4A     [Applications of Python](#) 838338005 #  
Session 4B     [Analysis using R-Shiny](#) 899402835 #  
Session 4C     [Lightning round](#) **Audio:** 1-866-865-9536, 744 124 3 #

# GASP 2019

## Government Advances in Statistical Programming workshop

Conference coordinator and program chair: Mike Jadoo (BLS)

Volunteer coordinator: Peter B. Meyer (BLS)

We thank many volunteers for their efforts.

Thanks again to sponsorship from the Washington Statistical Society, the Interagency R Users Group, and the FCSM Computational Statistics Interest Group.

Special thanks to:

