Government Advances in Statistical Programming (GASP)
Wednesday, 2 February – Thursday, 3 February 2022

Program

<table>
<tr><th>Time (ET)</th><th>Wednesday 2 February 2022</th></tr>
<tr><td>12:00 -12:27 pm</td><td>Welcome by Lisa Frehill, co-chair of the organizing committee

Nathan Cruze on CSPOS and FCSM

Carol Robbins of NSF, introduces Emilda Rivers

Emilda Rivers, Director, National Center for Science and Engineering Statistics, National Science Foundation
Carol Robbins of NCSES, thanks and handoff</td></tr>
<tr><td>12:30 – 1:45 pm</td><td>Paper session 1: Machine Learning and Natural Language Processing
Chair: Kelsey Gray

<table><tr><th>Title</th><th>Presenter</th></tr><tr><td>Improving Public Comment Review with Machine Learning and Advanced Analytics</td><td>Rob Chew (RTI International)</td></tr><tr><td>Bringing Efficiencies to Criminal Justice Manual Coding through Machine Learning</td><td>Peter Baumgartner (Explosion AI)</td></tr><tr><td>Machine-Learning Based Identification of Federally Funded Research Topics</td><td>Eric J. Oh (University of Virginia)</td></tr><tr><td>An Automated Data Collection and Record Linkage Pipeline for Program Evaluation</td><td>Brandon Sepulvado (NORC)</td></tr></table></td></tr>
<tr><td>1:45 – 2:30 pm</td><td>Lightning Talks 1: Machine Learning and Automation
Chair: Peter Baumgartner

<table><tr><th>Title</th><th>Presenter</th></tr><tr><td>Automating Checks for the CPS ASEC Poverty Detailed Tables</td><td>John Creamer (US Census Bureau)</td></tr><tr><td>Automating an ecological data workflow with GitHub Actions and R</td><td>Jeffrey W. Hollister (US Environmental Protection Agency)</td></tr><tr><td>imodels: a python package for fitting interpretable models</td><td>Chandan Singh (UC Berkeley)</td></tr><tr><td>h2o in R: machine learning in the Medical Expenditure Panel Survey</td><td>Emily Mitchell (AHRQ)</td></tr></table></td></tr>
<tr><td>2:30 – 2:45 pm</td><td>Break</td></tr>
</table>

<table>
<tr><th>Time (ET)</th><th>Wednesday 2 February 2022</th></tr>
<tr><td>2:45 – 3:30 pm</td><td>Lightning Talks 2: Natural Language Processing
Chair: Cecile Murray

<table><tr><th>Title</th><th>Presenter</th></tr><tr><td>Automating Quality Assessment of Survey Interviews</td><td>Kasey Jones (RTI International)</td></tr><tr><td>Using Natural Language Processing to Analyze Qualitative Data</td><td>Josh Cox (Regional Educational Laboratory Northeast & Islands)</td></tr><tr><td>ORS Task List Classification - Topic Modeling</td><td>Drake Gibson (Bureau of Labor Statistics)</td></tr><tr><td>Python for Paperwork Reduction Act Management and Reporting</td><td>Ed O'Brien (Department of Labor)</td></tr></table></td></tr>
<tr><td>3:30 – 5:00 pm</td><td>Paper session 2: Advances in Statistical Modeling
Chair: Drake Gibson

<table><tr><th>Title</th><th>Presenter</th></tr><tr><td>A Variable Selection Method for Small Area Estimation Modeling</td><td>Weijia Ren (Westat)</td></tr><tr><td>Regularization and Parametric Reduction for Fellegi-Sunter in Record Linkage</td><td>Daniel Weinberg (U.S. Census Bureau)</td></tr><tr><td>Bridging models to reconcile statistics based on data from multiple sources</td><td>Andreea Erciulescu (Westat)</td></tr><tr><td>A Spatiotemporal Analysis of County-Level Rental Rates</td><td>Wesley Burnett (Economic Research Service, USDA)</td></tr><tr><td>Two-Dimensional Statistical Manifold of Tornado Property Losses</td><td>Thilini Mahanama (Texas Tech University)</td></tr><tr><td>Wrap up of day 1: Drake Gibson</td><td></td></tr></table></td></tr>
</table>

<table>
<tr><th>Time (ET)</th><th>Thursday 3 February 2022</th></tr>
<tr><td>11:00 am – 12:00 pm</td><td>Cecile Murray's morning workshop:
Git Started with Version Control</td></tr>
<tr><td>12:00 - 12:15 pm</td><td>Day 2 welcome
Peter Meyer, co-chair welcome and logistics</td></tr>
<tr><td>12:15 – 1:00 pm</td><td>Lightning Talks 3: Development to Production
Chair: Kelsey Gray opens and at 12:15 pm and then at a little before 1:00 hands over to Christian Moscardi.

<table>
<tr><th>Title</th><th>Presenter</th></tr>
<tr><td>Paper, Prototype, Production: Application Development for Survey</td><td>Yong Li (Emily) Dich (US Census / Reveal Global Consulting)</td></tr>
<tr><td>Banff and Simputation: A Comparison Using BEA's Multinational Enterprise Surveys</td><td>Larkin Terrie (Bureau of Economic Analysis)</td></tr>
<tr><td>A System for Efficient Dataset Documentation Production</td><td>Weihuang Wong (NORC)</td></tr>
<tr><td>Modularized Statistical Disclosure Control In Data Warehouse Environment.</td><td>Russell Ballard (Accenture Federal Services)</td></tr>
<tr><td>"Moderation in Network Analysis on Time-Series Data Using R: A Case Study</td><td>Anastasia Galkina (UNIMIB)</td></tr>
</table></td></tr>
<tr><td>1:00 – 1:45 pm</td><td>Panel: Data Science's Real "Hidden Layer": IT Environments

Organizer and Chair: Christian Moscardi (U.S. Census Bureau)
Panelists:
Cecile Murray (U.S. Census Bureau)
Jess Klein (U.S. Census Bureau)
Taylor Hanson (U.S. Census Bureau)
Q&A between audience and panelists</td></tr>
<tr><td>1:45 – 2:00 pm</td><td>Break</td></tr>
</table>

<table>
<tr><th>Time (ET)</th><th>Thursday 3 February 2022</th></tr>
<tr><td>2:00 – 3:15 pm</td><td>Lightning Talks 4: Dashboards and Data Visualization
Chair: Carol Robbins

<table>
<tr><th>Title</th><th>Presenter</th></tr>
<tr><td>2020 Census Real Time Analysis of Data (RTAD)</td><td>Tracae McClure (U.S. Census Bureau)</td></tr>
<tr><td>America at Work: Multivariate data display for Labor Force Statistics [BLS]</td><td>Wendy B. Dickinson, Ph.D. (University of South Florida)</td></tr>
<tr><td>Building Web Dashboards with Harness-Vue</td><td>Alex Harding (RTI International)</td></tr>
<tr><td>Static to Interactive</td><td>Austin Lampros (Department of Transportation)</td></tr>
<tr><td>Creation of R Shiny Applications to Illustrate the Growclusters Package</td><td>Randall Powers (Bureau of Labor Statistics)</td></tr>
<tr><td>VeridicalFlow: building trustworthy data science pipelines with PCS</td><td>James Duncan (UC Berkeley)</td></tr>
<tr><td>“United States Marine Forces Special Operations Command (MARSOC) Preservation of the Force and Family POTFF Application Center: AI Automated HP System”</td><td>Joseph Lipoff (MARSOC)</td></tr>
</table></td></tr>
<tr><td>3:15 – 4:30 pm</td><td>Paper session 3: Data Systems, Usability, Reporting
Chair: Peter Meyer

<table>
<tr><th>Title</th><th>Presenter</th></tr>
<tr><td>Data Modernization for 'Health, US'</td><td>Barnali Das (National Center for Health Statistics)</td></tr>
<tr><td>Automated Collection of Publicly Available Data from the Internet</td><td>Mike Castro (U.S. Census Bureau)</td></tr>
<tr><td>Creating a performant & reliable estimation system using open-source tools</td><td>Ian Thomas (RTI International)</td></tr>
<tr><td>Helping Federal Clients Move from Table-based Reporting to Dynamic Dashboards</td><td>Marcia Underwood (RTI International)</td></tr>
</table></td></tr>
<tr><td>4:30 – 5:00 pm</td><td>Closing keynote presentation: Mike Linksvayer (GitHub)
Chair & interviewer: José (Bayoán) Santiago-Calderón (BEA)

Nathan Cruze: Closing remarks
Goodbye from all organizers</td></tr>
</table>

The organizers thank the NSF for providing the Zoom service for all attendees.
We thank Dahlia Saleh of NSF for helping organize the conference and being a contact point.
We thank Derrick Hampton of NSF for audio/visual support during the event.