

Program - Virtual Workshop

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

To Register: https://nsf.zoomgov.com/webinar/register/WN_Qu26asx0THitqbhK6MtxlA

| Time (ET) | Wednesday 2 February 2022 | | | | | | | | | | |
|--|---|-------|-----------|--|---------------------------------|--|--|---|---|---|----------------------------------|
| 12:00 – 12:30 pm | Welcome – Emilda Rivers, Director, National Center for Science and Engineering Statistics, National Science Foundation | | | | | | | | | | |
| 12:30 – 1:45 pm | 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>Anna Godwin (RTI International)</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> | Title | Presenter | Improving Public Comment Review with Machine Learning and Advanced Analytics | Rob Chew (RTI International) | Bringing Efficiencies to Criminal Justice Manual Coding through Machine Learning | Anna Godwin (RTI International) | Machine-Learning Based Identification of Federally Funded Research Topics | Eric J. Oh (University of Virginia) | An Automated Data Collection and Record Linkage Pipeline for Program Evaluation | Brandon Sepulvado (NORC) |
| Title | Presenter | | | | | | | | | | |
| Improving Public Comment Review with Machine Learning and Advanced Analytics | Rob Chew (RTI International) | | | | | | | | | | |
| Bringing Efficiencies to Criminal Justice Manual Coding through Machine Learning | Anna Godwin (RTI International) | | | | | | | | | | |
| Machine-Learning Based Identification of Federally Funded Research Topics | Eric J. Oh (University of Virginia) | | | | | | | | | | |
| An Automated Data Collection and Record Linkage Pipeline for Program Evaluation | Brandon Sepulvado (NORC) | | | | | | | | | | |
| 1:45 – 2:30 pm | 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> | Title | Presenter | Automating Checks for the CPS ASEC Poverty Detailed Tables | John Creamer (US Census Bureau) | Automating an ecological data workflow with GitHub Actions and R | Jeffrey W. Hollister (US Environmental Protection Agency) | imodels: a python package for fitting interpretable models | Chandan Singh (UC Berkeley) | h2o in R: machine learning in the Medical Expenditure Panel Survey | Emily Mitchell (AHRQ) |
| Title | Presenter | | | | | | | | | | |
| Automating Checks for the CPS ASEC Poverty Detailed Tables | John Creamer (US Census Bureau) | | | | | | | | | | |
| Automating an ecological data workflow with GitHub Actions and R | Jeffrey W. Hollister (US Environmental Protection Agency) | | | | | | | | | | |
| imodels: a python package for fitting interpretable models | Chandan Singh (UC Berkeley) | | | | | | | | | | |
| h2o in R: machine learning in the Medical Expenditure Panel Survey | Emily Mitchell (AHRQ) | | | | | | | | | | |
| 2:30 – 2:45 pm | Break | | | | | | | | | | |
| 2:45 – 3:30 pm | 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> | Title | Presenter | Automating Quality Assessment of Survey Interviews | Kasey Jones (RTI International) | Using Natural Language Processing to Analyze Qualitative Data | Josh Cox (Regional Educational Laboratory Northeast & Islands) | ORS Task List Classification - Topic Modeling | Drake Gibson (Bureau of Labor Statistics) | Python for Paperwork Reduction Act Management and Reporting | Ed O'Brien (Department of Labor) |
| Title | Presenter | | | | | | | | | | |
| Automating Quality Assessment of Survey Interviews | Kasey Jones (RTI International) | | | | | | | | | | |
| Using Natural Language Processing to Analyze Qualitative Data | Josh Cox (Regional Educational Laboratory Northeast & Islands) | | | | | | | | | | |
| ORS Task List Classification - Topic Modeling | Drake Gibson (Bureau of Labor Statistics) | | | | | | | | | | |
| Python for Paperwork Reduction Act Management and Reporting | Ed O'Brien (Department of Labor) | | | | | | | | | | |

| Time (ET) Wednesday 2 February 2022 | | | | | | | | | | | | | |
|--|--|-------|-----------|--|---------------------|--|--------------------------------------|---|-----------------------------|--|--|---|--|
| 3:30 – 5:00 pm | 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> </table> | Title | Presenter | A Variable Selection Method for Small Area Estimation Modeling | Weijia Ren (Westat) | Regularization and Parametric Reduction for Fellegi-Sunter in Record Linkage | Daniel Weinberg (U.S. Census Bureau) | Bridging models to reconcile statistics based on data from multiple sources | Andreea Erciulescu (Westat) | A Spatiotemporal Analysis of County-Level Rental Rates | Wesley Burnett (Economic Research Service, USDA) | Two-Dimensional Statistical Manifold of Tornado Property Losses | Thilini Mahanama (Texas Tech University) |
| Title | Presenter | | | | | | | | | | | | |
| A Variable Selection Method for Small Area Estimation Modeling | Weijia Ren (Westat) | | | | | | | | | | | | |
| Regularization and Parametric Reduction for Fellegi-Sunter in Record Linkage | Daniel Weinberg (U.S. Census Bureau) | | | | | | | | | | | | |
| Bridging models to reconcile statistics based on data from multiple sources | Andreea Erciulescu (Westat) | | | | | | | | | | | | |
| A Spatiotemporal Analysis of County-Level Rental Rates | Wesley Burnett (Economic Research Service, USDA) | | | | | | | | | | | | |
| Two-Dimensional Statistical Manifold of Tornado Property Losses | Thilini Mahanama (Texas Tech University) | | | | | | | | | | | | |

| Time (ET) Thursday 3 February 2022 | | | | | | | | | | | | | |
|--|--|-------|-----------|--|---|--|---|---|-----------------------|---|--|---|----------------------------|
| 11:00 am – 12:00 pm | Workshop: Git Started with Version Control with Cecile Murray | | | | | | | | | | | | |
| 12:00 – 12:15 pm | Day 2 welcome | | | | | | | | | | | | |
| 12:15 – 1:00 pm | Lightning Talks 3: Development to Production Chair: Kelsey Gray <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>Wei Huang 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> | Title | Presenter | Paper, Prototype, Production: Application Development for Survey | Yong Li (Emily) Dich (US Census / Reveal Global Consulting) | Banff and Simputation: A Comparison Using BEA's Multinational Enterprise Surveys | Larkin Terrie (Bureau of Economic Analysis) | A System for Efficient Dataset Documentation Production | Wei Huang Wong (NORC) | Modularized Statistical Disclosure Control In Data Warehouse Environment. | Russell Ballard (Accenture Federal Services) | "Moderation in Network Analysis on Time-Series Data Using R: A Case Study | Anastasia Galkina (UNIMIB) |
| Title | Presenter | | | | | | | | | | | | |
| Paper, Prototype, Production: Application Development for Survey | Yong Li (Emily) Dich (US Census / Reveal Global Consulting) | | | | | | | | | | | | |
| Banff and Simputation: A Comparison Using BEA's Multinational Enterprise Surveys | Larkin Terrie (Bureau of Economic Analysis) | | | | | | | | | | | | |
| A System for Efficient Dataset Documentation Production | Wei Huang Wong (NORC) | | | | | | | | | | | | |
| Modularized Statistical Disclosure Control In Data Warehouse Environment. | Russell Ballard (Accenture Federal Services) | | | | | | | | | | | | |
| "Moderation in Network Analysis on Time-Series Data Using R: A Case Study | Anastasia Galkina (UNIMIB) | | | | | | | | | | | | |
| 1:00 – 1:45 pm | Panel: Data Science's Real "Hidden Layer": IT Environments Organizer and Chair: Christian Moscardi (U.S. Census Bureau) Panelists: <ul style="list-style-type: none"> Cecile Murray (U.S. Census Bureau) Jess Klein (U.S. Census Bureau) Taylor Hanson (U.S. Census Bureau) | | | | | | | | | | | | |
| 1:45 – 2:00 pm | Break | | | | | | | | | | | | |

| Time (ET) | | Thursday 3 February 2022 | |
|---|---|---|--|
| 2:00 – 3:15 pm | Lightning Talks 4: Dashboards and Data Visualization | | |
| | Chair: Carol Robbins | | |
| | Title | Presenter | |
| | 2020 Census Real Time Analysis of Data (RTAD) | Tracae McClure (U.S. Census Bureau) | |
| | America at Work: Multivariate data display for Labor Force Statistics [BLS] | Wendy B. Dickinson, Ph.D. (University of South Florida) | |
| | Building Web Dashboards with Harness-Vue | Alex Harding (RTI International) | |
| | Static to Interactive: Successes and pitfalls when using publicly available government data to modernize the Department of Transportation's reporting | Austin Lampros (Department of Transportation) | |
| | Creation of R Shiny Applications to Illustrate the Growclusters Package | Randall Powers (Bureau of Labor Statistics) | |
| | VeridicalFlow: building trustworthy data science pipelines with PCS | James Duncan (UC Berkeley) | |
| 3:15 – 4:30 pm | "United States Marine Forces Special Operations Command (MARSOC) Preservation of the Force and Family POTFF Application Center: AI Automated HP System" | | |
| | Paper session 3: Data Systems, Usability, Reporting | | |
| | Chair: Peter Meyer | | |
| | Title | Presenter | |
| | Data Modernization for 'Health, US' | Barnali Das (National Center for Health Statistics) | |
| 4:30 – 5:00 pm | Automated Collection of Publicly Available Data from the Internet | Mike Castro (U.S. Census Bureau) | |
| | Creating a performant & reliable estimation system using open-source tools | Ian Thomas (RTI International) | |
| | Helping Federal Clients Move from Table-based Reporting to Dynamic Dashboards | Marcia Underwood (RTI International) | |
| | Closing remarks: Mike Linksvayer (GitHub) | | |
| Chair: José (Bayoán) Santiago-Calderón | | | |

Presented by the Computational Statistics for the Production of Official Statistics (CSPOS) Interest Group of the Federal Committee on Statistical Methods (FCSM).