

Printable Program  
Printed on: 12/09/2018

Wednesday, May 16

Registration  
SDSS Hours  
Wed, May 16, 7:00 AM - 6:30 PM  
Registration

SC1 - Data Science Workflows Using R and Spark  
Short Course  
Wed, May 16, 8:00 AM - 5:30 PM  
Grand Ballroom E

*Instructor(s): Jim Harner, West Virginia University and NISS*

This short course covers the data science process using R as a programming language and Spark as a big-data platform. Powerful workflows are developed using the `tidyr`, `dplyr`, `ggplot2`, and `sparklyr` packages. Examples show how data is transported to and extracted from persistent data stores such as the Hadoop Distributed File System (HDFS), NoSQL databases, and relational databases. These data-based workflows extend to machine learning algorithms, model evaluation, and data visualization. TensorFlow for deep learning is introduced. Big-data architectures are discussed including the Docker containers used for building the course infrastructure called `rspark` (<https://github.com/jharner/rspark>). Attendees can optionally install Docker containers on their desktop or deploy them to Amazon Web Services (AWS) prior to the course (see the `rspark` repo).

SC2 - H2O AutoML  
Short Course  
Wed, May 16, 8:00 AM - 12:00 PM  
Grand Ballroom G

*Instructor(s): Navdeep Gill, H2O.ai*

In recent years, the demand for machine learning experts has outpaced the supply, despite the surge of people entering the field. To address this gap, there have been big strides in the development of user-friendly machine learning software that can be used by non-experts. Although H2O has made it easier for practitioners to train and deploy machine learning models at scale, there is still a fair bit of knowledge and background in data science that is required to produce high-performing machine learning models. Deep Neural Networks in particular, are notoriously difficult for a non-expert to tune properly. In this course, we provide an overview of the the field of "Automatic Machine Learning" and introduce the new AutoML functionality in H2O. H2O's AutoML provides an easy-to-use interface which automates the process of training a large, comprehensive selection of

candidate models and a stacked ensemble model which, in most cases, will be the top performing model in the AutoML Leaderboard. H2O AutoML is available in all the H2O interfaces including the h2o R package, Python module and the Flow web GUI. We will also provide code examples to get you started using AutoML.

SC3 - End-to-End Machine Learning and Model Deployment in SAS® Viya®  
Short Course  
Wed, May 16, 8:00 AM - 12:00 PM  
Lake Fairfax A

*Instructor(s): Carlos Pinheiro, SAS & Data Science Tech Institute, France*

In this course, you will learn the latest groundbreaking interface from SAS® which uses a pipeline flow approach to:

- access, manage, and explore data
- develop and compare models
- generate and register score code
- publish champion models in a database
- export score code to files

SC4 - CANCELLED: Cloudera Data Science Workbench (CDSW)  
Short Course  
Wed, May 16, 1:30 PM - 5:30 PM

*Instructor(s): TBD TBD, Cloudera*

Join your peers at a Cloudera hosted short course to discuss your Data Science needs across your organization. Machine learning and Data Science are all about the data, but it's often out of reach for analytics teams working at scale.

Together we'll explore how to leverage powerful open source tools to create a machine learning mixture that balances data scientists' need for data access and flexible tooling with IT needs for security and governance. Cloudera Data Science Workbench enables fast, easy, and secure self-service data science in a collaborative environment.

Ultimately you'll walk away prepared to discover a new way to find value in your data and deliver increased value to your organization.

SC5 - Shiny Essentials  
Short Course  
Wed, May 16, 1:30 PM - 5:30 PM  
Grand Ballroom G

*Instructor(s): Mine Cetinkaya-Rundel, Duke University & RStudio*

Shiny is an R package that makes it easy to build interactive web apps straight from R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. This short course will introduce you to the basics of building web applications with Shiny, essentials of reactive programming, and how to customize and deploy your apps for others to use. Please bring a laptop with you to the course.

Exhibits Open  
SDSS Hours  
Wed, May 16, 5:30 PM - 7:00 PM  
Regency Ballroom Foyer

PS01 - Opening Mixer & General E-Posters  
E-Poster  
Wed, May 16, 5:30 PM - 7:00 PM  
Regency Ballroom

E-Poster session will take place from 5:45 p.m. - 6:45 p.m.

- 1  
[Box Plots and Q-Q Plots for Geometric and Harmonic Observations](#)  
*Mian Arif Shams Adnan, Indiana University Bloomington*
- 2  
[Changepoint Detection Using Subsampling and Knockoff Variables](#)  
*Sangwon Hyun, Department of Statistics, Carnegie Mellon University*
- 4  
[Modified Wald Test for Reference Scaled Equivalence Assessment of Analytical Biosimilarity](#)  
*Yu-Ting Weng, FDA*
- 5  
[Incremental Parameter Estimation for a Massively Multi-Parameter Regression Model](#)  
*David I. Donato, U.S. Geological Survey*
- 6  
[Data Issues in Modeling and Estimation of Urban Transportation Networks](#)  
*Isabelle Kemajou-Brown, Morgan State University*

7

[Machine Learning Methods for Animal Movement](#)

*Dhanushi A Wijeyakulasuriya, Pennsylvania State University*

8

[Using Software to Quantify Estimation Uncertainty in Statistical Results](#)

*Jordan Lee Prendez, University of Maryland*

Thursday, May 17

Registration

SDSS Hours

Thu, May 17, 7:30 AM - 5:30 PM

Registration

Exhibits Open

SDSS Hours

Thu, May 17, 7:30 AM - 7:15 PM

Regency Ballroom Foyer

GS01 - Welcome and Keynote Address

General Session

Thu, May 17, 8:30 AM - 10:00 AM

Grand Ballroom D

*Chair(s): Yasmin H. Said, George Mason University*

8:30 AM

[SDSS Welcome](#)

*Yasmin H. Said, George Mason University; Jim Harner, West Virginia University; Ronald L. Wasserstein, American Statistical Association*

8:45 AM

[Uncovering the Mechanisms of General Anesthesia: Where Neuroscience Meets Statistics](#)

*Emery N. Brown, MIT, Harvard Medical School, and Massachusetts General Hospital*

9:40 AM

[Edward J. Wegman Award Ceremony](#)

PS02 - Public Health/Disease

E-Poster

Thu, May 17, 10:00 AM - 10:45 AM

Regency Ballroom B

1

[Daily Smokers' Attributes Associated with Purchasing Cigarettes on Indian Reservations](#)

*Richard A Pack, Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida*

2

[120/5000 Estimation of Life Years Potentially Lost Due to Traffic Accidents Involving a Motorcycle in Costa Rica](#)

*Agustín Gómez Meléndez, University of Costa Rica*

3

[Mapping Rates of Inpatient Hospitalizations Related to Mental Disorders in the State of Missouri: A Conditional Autoregressive Model With Zip Code-Level Data](#)

*Daphne Lew, Saint Louis University*

4

[Improved Predictive Models for Readmission of Patients with Diabetes](#)

*Chathurangi Heshani Karunapala Pathiravasan, Southern Illinois University*

5

[Development of Prognostic Model for Breast Cancer in Shanghai Breast Cancer Survival Study \(SBCSS\)](#)

*Run Fan, Vanderbilt University Medical Center, Department of Biostatistics*

6

[A Machine Learning Approach to Improve Fall Risk Prediction in Home Health Care](#)

*Yancy Lo, Institute for Biomedical Informatics, The Perelman School of Medicine, University of Pennsylvania*

7

[Treating Leukemia in Youths](#)

*Zachary R Smith, University of Michigan - Dearborn*

8

[Survival of Young Leukemia Patients](#)

*Theren Williams, University of Michigan- Dearborn*

9

[Hospital Readmission Risk Prediction after Joint Replacement Surgery](#)

*Selah F. Lynch, Institute for Biomedical Informatics, The Perelman School of Medicine, University of Pennsylvania*



CS01 - Automated Model Building

Invited

Thu, May 17, 10:30 AM - 12:00 PM

Grand Ballroom D

*Organizer(s): William S. Cleveland, Purdue*

*Chair(s): Ryan Hafen, Hafen Consulting LLC*

10:30 AM

[D3M Automated Model Building and Diagnostics](#)

*Curtis Lisle, KnowledgeVis LLC*

11:00 AM

[Candela: An Interactive Visualization Component Library for Data Science](#)

*Jeffrey Baumes, Kitware, Inc.*

11:30 AM

[Average-Transform-Smooth \(ATS\) Diagnostic Methods for Non-Gaussian Models](#)

*William S. Cleveland, Purdue*



CS02 - Statistics Inference for High-Dimensional Regression

Invited

Thu, May 17, 10:30 AM - 12:00 PM

Grand Ballroom E

*Organizer(s): Larry Wasserman, Carnegie Mellon University*

*Chair(s): Todd A Kuffner, Washington University in St. Louis*

10:30 AM

[Testing for Global Network Structure Using Small Subgraph Statistics](#)

*Chao Gao, University of Chicago*

11:00 AM

[Inferential Goals, Targets, and Principles in High-Dimensional Regression](#)

*Todd A Kuffner, Washington University in St. Louis*

11:30 AM

[Selective Inference in Linear Regression](#)

*Jonathan Taylor, Stanford University*



CS03 - Interactive Statistical Graphics: Where Are We Now?

Invited

Thu, May 17, 10:30 AM - 12:00 PM

Grand Ballroom F

*Organizer(s): Adalbert Wilhelm, Jacobs University*

*Chair(s): Adalbert Wilhelm, Jacobs University*

10:30 AM

[Exploratory Visualization via Extendible Interactive Graphics](#)

*Wayne Oldford, University of Waterloo*

11:00 AM

[Model Exploration via Conditional Visualisation](#)

*Catherine Hurley, Maynooth University*

11:30 AM

[Interactive \(Web-\)Graphics \(using R\)](#)

*Heike Hofmann, Iowa State University*



CS04 - Best Practices in Data Science Education

Invited

Thu, May 17, 10:30 AM - 12:00 PM  
Grand Ballroom G

*Organizer(s): Ben Baumer, Smith College*

*Chair(s): Ben Baumer, Smith College*

10:30 AM

[Start with Data Science as an Introduction to Statistical Thinking](#)

*Mine Cetinkaya-Rundel, Duke University & RStudio*

11:00 AM

[Data Science for Everybody: Building and Characterizing Student-Driven Pathways in Introductory Statistics Courses](#)

*Rebecca Nugent, Carnegie Mellon Statistics & Data Science*

11:30 AM

[Data-Driven Curriculum Development](#)

*David Robinson, DataCamp*



CS05 - Statistical Machine Learning with Business Applications

Invited

Thu, May 17, 10:30 AM - 12:00 PM

Regency Ballroom A

*Organizer(s): Brad Price, West Virginia University*

*Chair(s): Brad Price, West Virginia University*

10:30 AM

[A Cluster Elastic Net for Multivariate Regression](#)

*Ben Sherwood, University of Kansas*

11:00 AM

[Selection and Its Inference Using the Whole Solution Paths](#)

*Peng Wang, University of Cincinnati*

11:30 AM

[Shrinking Characteristics of Precision Matrix Estimators](#)

*Aaron J. Molstad, Fred Hutchinson Cancer Research Center*



CS06 - Analytics for Fitness Tracker Data

Invited

Thu, May 17, 10:30 AM - 12:00 PM

Lake Fairfax A

*Organizer(s): David Marchette, Naval Surface Warfare Center*

*Chair(s): Shelby Macy, Naval Surface Warfare Center*

10:30 AM

[Correlating Sleep and Temperature Patterns in Navy Warfighters With Current and Future Health Status](#)

*Laura Maple, NSWCDD*

11:00 AM

[An Artificial Intelligence System for Real-Time Individualized Core Temperature Estimation](#)

*Jaques Reifman, US Army MRMC/BHSAI*

11:30 AM

[Statistical Methods for Micro- and Macro-Level Accelerometry Data](#)

*Jiawei Bai, Johns Hopkins University*



CS07 - Optimization

Contributed

Thu, May 17, 10:30 AM - 12:00 PM

Lake Fairfax B

*Chair(s): Jingyi Zhu, The Johns Hopkins University*

10:30 AM

[Topological Mixture Estimation](#)

*Steve Huntsman, BAE Systems*

10:45 AM

[Plotting Two-Dimensional Confidence Regions](#)

*Christopher Weld, College of William & Mary*

11:00 AM

[Tracking Capability of Stochastic Approximation Algorithms with Constant Gain](#)

*Jingyi Zhu, The Johns Hopkins University*

11:15 AM

[Variable Selection for Consistent Clustering](#)

*Ronald Joseph Yurko, Carnegie Mellon University*

11:30 AM

[BRISC: Bootstrap for Rapid Inference on Spatial Covariances](#)

*Arkajyoti Saha, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

11:45 AM

[Reduced Complexity of Second-Order Simultaneous Perturbation Stochastic Approximation Algorithms](#)

*Jingyi Zhu, The Johns Hopkins University*



CS08 - Reasoning with Data

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Grand Ballroom D



*Organizer(s): William Szewczyk, Mathematics Research Group, National Security Agency*

*Chair(s): William Szewczyk, Mathematics Research Group, National Security Agency*

1:30 PM

[Capturing Subject Matter Expertise for Automated Assisted Analysis](#)

*William Szewczyk, Mathematics Research Group, National Security Agency*

2:00 PM

[Task-Centric Document Curation based on Node Embeddings from a Graphical Representation of Workflows](#)

*Paul Jones, Laboratory for Analytic Sciences*

2:30 PM

[Experiences with AI, Expert Knowledge and Data Analysis](#)

*Octavian Udrea, IBM T.J. Watson Research Center*



CS09 - Advanced Mathematics for Data Analysis

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Grand Ballroom E

*Organizer(s): David Marchette, Naval Surface Warfare Center*

*Chair(s): David Marchette, Naval Surface Warfare Center*

1:30 PM

[Persistence Images and Applications](#)

*Tegan Emerson, Naval Research Laboratory*

2:00 PM

[A Geometric Formulation of Neural Network Training](#)

*David A. Johannsen, Naval Surface Warfare Center - Dahlgren*

2:30 PM

[Information Tests on Statistical Submanifolds](#)

*Michael Trosset, Indiana University*



CS10 - Visualization Using Open-Source Tools

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Grand Ballroom F

*Organizer(s): Wendy Martinez, U.S. Bureau of Labor Statistics*

*Chair(s): Wendy Martinez, U.S. Bureau of Labor Statistics*

1:30 PM

[Visualizing BLS Data in Google Public Data Explorer](#)

*Christopher Morris, U.S. Bureau of Labor Statistics*

2:00 PM

[Visualization Using Open-Source Tools: some FDA perspectives](#)

*Paul Schuette, US Food and Drug Administration*

2:30 PM

[Small Business Database](#)

*Richard Schwinn, Small Business Administration*



CS11 - Big Data Analytics Using R and Spark

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Grand Ballroom G

*Organizer(s): Brad Price, West Virginia University*

*Chair(s): Brad Price, West Virginia University*

1:30 PM

[Data Science Workflows](#)

*Jim Harner, West Virginia University*

2:00 PM

[Data Science at Scale With R and Sparklyr: Architecture, Ecosystem, and Current Developments](#)

*Kevin Kuo, Rstudio*

2:30 PM

[Interacting with Distributed Data from R using SparkR](#)

*Hossein Falaki, Databricks*



CS12 - Model Selection in High-Dimensions with Complexities

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Regency Ballroom A

*Organizer(s): Hamparsum Bozdogan, University of Tennessee*

*Chair(s): Hamparsum Bozdogan, University of Tennessee*

1:30 PM

[A New Approach to Dimension Reduction For Multivariate Time Series](#)

*Chung Eun Lee, University of Tennessee, Knoxville*

2:00 PM

[Coordinate-Independent Sparse Estimation in Semiparametric Models](#)

*Haileab Hilafu, University of Tennessee*

2:30 PM

[Expected Volume Confidence Region Complexity \(EVCR\\_COMP\) Criterion in High Dimensions](#)

[with Applications](#)

*Hamparsum Bozdogan, University of Tennessee*



CS13 - Social Network Analysis

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Lake Fairfax A

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): William F. Wiecek, SUNY Buffalo State*

1:30 PM

[Social Networks and Simplicial Complexes](#)

*Daniele Struppa, Chapman University*

2:00 PM

[Reflections on Computational Social Science, in Honor of Ed Wegman](#)

*Claudio Cioffi-Revilla, George Mason University*

2:30 PM

[The Big Picture: Big Data, Big Theory, and Big Challenges](#)

*William G. Kennedy, George Mason University*



CS14 - Monitoring Financial Stability with Data Science

Invited

Thu, May 17, 1:30 PM - 3:00 PM

Lake Fairfax B

*Organizer(s): Shawn Mankad, Cornell University*

*Chair(s): Shawn Mankad, Cornell University*

1:30 PM

[Modeling and Prediction of Financial Trading Networks: An Application to the NYMEX Natural Gas Futures Market](#)

*Abel Rodriguez, University of California, Santa Cruz*

2:00 PM

[Elicitability and Backtesting: Perspectives for Banking Regulation](#)

*Natalia Nolde, University of British Columbia*

2:30 PM

[Systemic Risk from Asset Concentration and Common Holdings among Banks](#)

*Celso Brunetti, Federal Reserve Board*

PS03 - Bayesian Modeling

E-Poster

Thu, May 17, 3:00 PM - 3:45 PM

1

[Bayesian Modeling of Non-Stationary, Univariate, Spatial Data](#)

*Margaret Goldman, U.S. Geological Survey*

2

[Choosing Among a Class of Zellner's g-Priors in Bayesian Regression Models and Subset Selection of Variables Using the Genetic Algorithm and Information Complexity](#)

*Yaojin Sun, The University of Tennessee*

3

[Lagged Exact Bayesian Online Changepoint Detection](#)

*Michael Byrd, Southern Methodist University*

4

[Constrained Bayesian Inference through Posterior Projections](#)

*Sayan Patra, Duke University*

5

[On the Quantification and Efficient Propagation of Imprecise Probabilities Using Monte Carlo Methods](#)

*Jiaxin Zhang, Johns Hopkins University*

6

[Bayesian Optimization of Personalized Models for Real-Time Patient Monitoring](#)

*Glen Wright Colopy, Oxford University*



CS15 - Best of Computational and Graphical Statistics

Invited

Thu, May 17, 3:30 PM - 5:00 PM

Grand Ballroom E

*Organizer(s): Di Cook, Monash University*

*Chair(s): Catherine Hurley, Maynooth University*

3:30 PM

[Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics](#)

*Susan Ruth VanderPlas, Nebraska Public Power District*

4:00 PM

[Fused Lasso Additive Model](#)

*Ashley Petersen, Division of Biostatistics, University of Minnesota*

4:30 PM

[Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE](#)

*Daniel Turek, Williams College*



## CS16 - Text Data Analytics and Visualization

Invited

Thu, May 17, 3:30 PM - 5:00 PM

Grand Ballroom E

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Kelly S Marczynski, SUNY Buffalo State*

3:30 PM

[Algorithmic and Visualization Frameworks to Facilitate the Revelation of Interesting Structure in Document Collections](#)

*Jeffrey L. Solka, Naval Surface Warfare Center*

4:00 PM

[Fast k Nearest Neighbor Graph Construction Experiments on a Large Scientific Publication Corpus](#)

*Avory Bryant, Naval Surface Warfare Center*

4:30 PM

[Leveraging Automated Storytelling With b-Privy Analytics: Creating Plausible Explanations of Emerging Technologies to Mitigate Surprise](#)

*John T. Rigsby, Naval Surface Warfare Center*



## CS17 - Data Science at the National Institute of Statistical Sciences

Invited

Thu, May 17, 3:30 PM - 5:00 PM

Grand Ballroom G

*Organizer(s): Jim Rosenberger, NISS and Pennsylvania State University*

*Chair(s): Jim Rosenberger, NISS and Pennsylvania State University*

3:30 PM

[Using Administrative Data to Produce Official Statistics: An Application to End-Of-Season Acreage Estimation](#)

*Andreea L Erciulescu, National Institute of Statistical Sciences and USDA National Agricultural Statistics Service*

4:00 PM

[Future of Integer Calibration Weighting Methods](#)

*Luca Sartore, National Institute of Statistical Sciences*

4:30 PM

[The NCES/NISS Partnership: Data Collection Efforts/Structures/New Initiatives](#)

*Nell Sedransk, National Institute of Statistical Sciences*



## CS18 - Nonlinear Dimension Reduction

Invited

Thu, May 17, 3:30 PM - 5:00 PM

## Regency Ballroom A

*Organizer(s): Michael Trosset, Indiana University*

*Chair(s): Michael Trosset, Indiana University*

3:30 PM

[Optimality of the Johnson-Lindenstrauss Lemma](#)

*Jelani Nelson, Harvard University*

4:00 PM

[Matrix Sketching for Alternating Direction Method of Moments Optimization](#)

*Daniel McDonald, Indiana University*

4:30 PM

[Optimal Dimensionality Reduction for Non-Linear Clustering Via Nystrom Approximation](#)

*Alex Gittens, Rensselaer Polytechnic Institute*



CS19 - CyberLanguage: Applications of Natural Language Processing to CyberSecurity  
Invited

Thu, May 17, 3:30 PM - 5:00 PM

Lake Fairfax A

*Organizer(s): Joseph Marr, DZYNE Technologies*

*Chair(s): Joseph Marr, DZYNE Technologies*

3:30 PM

[Network Traffic Anomaly Detection Using Recurrent Neural Networks](#)

*Benjamin Radford, KeyW*

4:00 PM

[Modeling Machine-to-Machine Cyber Data as Discrete Sequences of Activity](#)

*Bartley Richardson, KeyW*

4:30 PM

[Time Series Pattern Mining and Visualization Using Statistical Language Processing Techniques](#)

*Jessica Lin, George Mason University*



CS20 - Differential and Bitcoin Privacy

Invited

Thu, May 17, 3:30 PM - 5:00 PM

Lake Fairfax B

*Organizer(s): Roy E. Welsch, MIT*

*Chair(s): Roy E. Welsch, MIT*

3:30 PM

[Differentially Private Model Selection with Penalized and Constrained Likelihood](#)

*Jing Lei, Carnegie Mellon University*

4:00 PM

[Blockchain Technology: A New Approach to Digital Privacy?](#)

*Christian Catalini, MIT*

4:30 PM

[Differentially Private Parametric Inference](#)

*Marco Avella Medina, MIT*



CS21 - Computational Text Processing

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Grand Ballroom E

*Organizer(s): Mark Hansen, Columbia*

*Chair(s): Mark Hansen, Columbia*

5:15 PM

[Modeling and Understanding Language with Neural Networks Using Spark and R](#)

*Ali Zaidi, Microsoft AI and Research*

5:45 PM

[Computational Propaganda](#)

*Mark Hansen, Columbia*



CS22 - Distinguished Colleagues of Edward Wegman: Mathematical Physics

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Grand Ballroom F

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): David Marchette, Naval Surface Warfare Center*

5:15 PM

[Laws of the Universe, Information and Mind in the Quantum Universe](#)

*Menas C. Kafatos, Chapman University*

5:45 PM

[Exploring and Exploiting Interestingness in Data Science](#)

*Kirk Borne, Booz Allen Hamilton*



CS23 - Data Science Platforms I

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Grand Ballroom G

*Organizer(s): Jim Harner, West Virginia University*

*Chair(s): Jim Harner, West Virginia University*

5:15 PM

[Automating Data Science Processes with H2O Driverless AI](#)

*Patrick Hall, H2O.ai*

5:45 PM

[Building Data Science Platforms Using Docker](#)

*Jim Harner, West Virginia University*



CS24 - TensorFlow

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Regency Ballroom A

*Organizer(s): Tim Hesterberg, Google*

*Chair(s): Tim Hesterberg, Google*

5:15 PM

[TensorFlow Autograph: Source Code Transformation for Easier TensorFlow](#)

*Alex Wiltschko, Google*

5:45 PM

[Machine Learning with TensorFlow and R](#)

*J.J. Allaire, Rstudio*



CS25 - Time Series Modeling

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Lake Fairfax A

*Organizer(s): Jim Harner, West Virginia University*

*Chair(s): Rida Moustafa, Walmart*

5:15 PM

[The Divergence Between Observed and Modeled Temperature Trends in the Tropical Troposphere 1958-2017](#)

*Ross McKittrick, University of Guelph*

5:45 PM



[Forecasting with Many Predictors](#)

*Kyle Caudle, SD School of Mines and Technology*



CS26 - Combining Federal and Regional Data Sources: Challenges and Solutions

Invited

Thu, May 17, 5:15 PM - 6:15 PM

Lake Fairfax B

*Organizer(s): Lingzhou Xue, Pennsylvania State University*

*Chair(s): Nell Sedransk, National Institute of Statistical Sciences*

5:15 PM

[Six Classes of Methodological Research Questions in the Integration of Multiple Data Sources for Granular Estimation](#)

*John Eltinge, U.S. Census Bureau*

5:35 PM

[Use of the Quarterly Census of Employment and Wages and Third-Party Sources for EIA Surveys](#)

*Nanda Srinivasan, Energy Information Administration*

5:55 PM

[Discussant](#)

*Jim Rosenberger, NISS and Pennsylvania State University*

PS04 - Machine Learning Applications

E-Poster

Thu, May 17, 6:15 PM - 7:15 PM

Regency Ballroom B

1

[Penalized Regression Within the Game Cribbage](#)

*Christopher Silberstein, The Ohio State University*

2

[Diagnosing and Predicting the Eyewall Replacement Cycle: Learning from Hurricane Irma](#)

*Martha Lisbeth Christino, T.C. Williams High School*

3

[Random Forest Prediction Intervals](#)

*Haozhe Zhang, Iowa State University*

4

[Machine Learning for Acute Kidney Injury with IDEAs: Intraoperative Data Embedded Analytics](#)

*Lasith Adhikari, University of Florida*

5

[Predicting Human Alteration of River and Stream Salinity Using Random Forest Models](#)

*Franco Alexis Sanchez, California State University, Monterey Bay, Department of Mathematics and Statistics*

6

[Performance of Cross-Validation of Binary Longitudinal Finite Mixture Models: A Simulation and](#)

[Application.](#)

*Thom J Taylor, Nicklaus Childrens Research Institute*

7

[The Sliding Window Fourier Transform](#)

*Lee F Richardson, Carnegie Mellon university*

8

[Machine Learning Improved Classification of Psychoses using Clinical and Biological stratification: Update from the Bipolar-Schizophrenia Network for Intermediate Phenotypes \(B-SNIP\)](#)

*Suraj Sarvode Mothi, Department of Psychiatry, Massachusetts General Hospital*

9

[Inter- and Intra-Institutional Efforts to Build Capacity for Data Science Education](#)

*Douglas Landsittel, University of Pittsburgh*

GS02 - Symposium on Data Science & Statistics Banquet

General Session

Thu, May 17, 7:15 PM - 8:30 PM

Grand Ballroom D

[I Never Met a Datum I Didn't Like](#)

*Barry D. Nussbaum, 2017 President, American Statistical Association*

Friday, May 18

Exhibits Open

SDSS Hours

Fri, May 18, 7:30 AM - 4:00 PM

Regency Ballroom Foyer

Registration

SDSS Hours

Fri, May 18, 7:30 AM - 5:30 PM

Registration

GS03 - Plenary Session: Contributions to Computational Statistics

General Session

Fri, May 18, 8:30 AM - 10:00 AM

Grand Ballroom D

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Yasmin H. Said, George Mason University*

8:35 AM

[Ed Wegman's Influence on the Profession: His Work in Computational Statistics and Density](#)

[Estimation in Particular](#)

*David Scott, Rice University*

9:00 AM

[Statistical Graphics in Data Science](#)

*Adalbert Wilhelm, Jacobs University*

9:25 AM

[Omnibus Regression: Predicting Probability Distributions with Imperfect Data](#)

*Jerome H. Friedman, Stanford University*

9:50 AM

[Floor Discussion](#)

PS05 - Bioinformatics/Biomedical

E-Poster

Fri, May 18, 10:00 AM - 10:45 AM

Regency Ballroom B

1

[Effect of Non-Parametric Mapping Over Parametric Mapping for fMRI](#)

*Siddharth Nayak, Institute of Statistical Science, Academia Sinica*

2

[Identifying Bioethical Issues in Biostatistical Consulting: Findings From a US National Pilot Survey of Biostatisticians](#)

*Min Qi Wang, University of Maryland*

3

[Diagnostic Prediction of Autism in Resting-State Functional Mri Using Conditional Random Forest](#)

*Afroz Jahedi, San Diego State University*

4

[Data-Driven Statistical Methods for Detecting Gait Instability Using Physiological Signal Metrics](#)

*Kristin Morgan, University of Connecticut*

5

[A Comparison of Selected Parametric and Non-Parametric Statistical Approaches for Candidate Genes Selection in Transcriptome Data](#)

*Dawit Gezahegn Tadesse, Cincinnati Children's Hospital Medical Center*

6

[Wavelet-based Classification Applied to fMRI](#)

*Pedro Alberto Morettin, University of São Paulo*

7

[Visualizations to Guide Dimension Reduction for Sparse High-Dimensional Data](#)

*Snehalata Huzurbazar, West Virginia University*



CS27 - Distinguished Colleagues of Edward Wegman: Applications to Data Science

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Grand Ballroom D

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Yasmin H. Said, George Mason University*

10:30 AM

[Automatic Visualization](#)

*Leland Wilkinson, H2O*

11:00 AM

[Cherry-Picking for Complex Datasets](#)

*David Banks, SAMSI and Duke University*

11:30 AM

[Bayesian Penalty Mixing with the The Spike and Slab Lasso](#)

*Edward George, University of Pennsylvania*



CS28 - Bayesian Computations and Applications

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Grand Ballroom E

*Organizer(s): Ehsanolah Soofi, University of Wisconsin at Milwaukee*

*Chair(s): Ehsanolah Soofi, University of Wisconsin at Milwaukee*

10:30 AM

[Analysis of Crimean-Congo Hemorrhagic Fever Incidents with Dynamically Weighted Particle Filter](#)

*Duchwan Ryu, Northern Illinois University*

11:00 AM

[Non-Negative Matrix Factorization for The Exponential Family Based on Generalized Dual Divergence and Intrinsic Information](#)

*Karthik Devarajan, Fox Chase Cancer Center, Temple University Health System*

11:30 AM

[Masking Data Using an Entropy Approach](#)

*Kurt Pflughoeft, University of Wisconsin Milwaukee*



CS29 - Big Data Visualization

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Grand Ballroom F

*Organizer(s): Rida Moustafa, Walmart*

*Chair(s): Rida Moustafa, Walmart*

10:30 AM

[Developing Inferential Visual Analytics Systems for Scientific Applications](#)

*Chad A. Steed, Oak Ridge National Laboratory*

11:00 AM

[Data Visualization in Statistical Consulting Applications](#)

*Heather Watson, Exponent, Inc.*

11:30 AM

[Quantization and Enveloping Methods for Scaling Visualization Techniques to Big Data](#)

*Rida Moustafa, Walmart*



CS30 - Data Science Programs

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Grand Ballroom G

*Organizer(s): Tim Hesterberg, Google*

*Chair(s): Tim Hesterberg, Google*

10:30 AM

[NYU Master of Science in Data Science](#)

*Arthur Spirling, New York University*

10:55 AM

[Columbia University Master of Science in Data Science](#)

*Tian Zheng, Columbia University*

11:20 AM

[WVU Master of Science in Business Data Analytics: Challenges and Experiences with Online Data Science Programs](#)

*Brad Price, West Virginia University*

11:45 AM

[Floor Discussion](#)



CS31 - Recent Advances in Statistical Machine Learning

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Regency Ballroom A

*Organizer(s): Eric Chi, North Carolina State University; David Scott, Rice University*

*Chair(s): David Scott, Rice University*

10:30 AM

[On the Regularizations for Enforcing Equi-Sparsity](#)

*Yiyuan She, Florida State University*

11:00 AM

[An Alternating Directions Method for Large-scale Multivariate Convex Regression](#)

*Jason Xu, University of California Los Angeles*

11:30 AM

[Tensor Canonical Correlation Analysis](#)

*Eric Chi, North Carolina State University*



CS32 - Data Science Partnerships

Invited

Fri, May 18, 10:30 AM - 12:00 PM

Lake Fairfax A

*Organizer(s): Sallie Keller, Biocomplexity Institute of Virginia Tech*

*Chair(s): Sallie Keller, Biocomplexity Institute of Virginia Tech*

10:30 AM

[Using Multiple Big Data Sources to Manage a Supply Chain](#)

*Dave Higdon, SDAL, Virginia Tech*

11:00 AM

[Partnering for Data Science: The Laboratory for Analytic Sciences](#)

*Alyson Wilson, North Carolina State University*

11:30 AM

[University, Government, NGO Partnership Around Statistical Solutions to Urban Challenges](#)

*Katherine Bennett Ensor, Rice University*



CS33 - Survey Science

Contributed

Fri, May 18, 10:30 AM - 12:00 PM

Lake Fairfax B

*Chair(s): MoonJung Cho, U.S. Bureau of Labor Statistics*

10:30 AM

[Survey Estimation with Elastic Net Regression: Combining Data Sources to Improve Estimator Efficiency](#)

*Kelly Sue McConville, Swarthmore College*

10:45 AM

[Pseudolikelihood Inference for Quantiles From Complex Surveys](#)

*Jing Wang, The University of Texas at Arlington*

11:00 AM

[Can a Statistician Thrive Using Only Free Software?](#)

*Amang Sukasih, RTI International*

11:15 AM

[Systematic Sampling Design with Application to Data Splitting](#)

*Redouane Betrouni, George Mason University*

11:30 AM

[Incorporating Design Concepts and Methods into the Integration of Multiple Data Sources](#)

*John Eltinge, U.S. Census Bureau*

11:45 AM

[Classification Trees for Privacy in Sample Surveys](#)

*Rolando Andres Rodriguez, U.S. Census Bureau*



CS34 - Distinguished Students of Edward Wegman

Invited

Fri, May 18, 1:30 PM - 3:00 PM

Grand Ballroom D

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Edward George, University of Pennsylvania*

1:30 PM

[On Spectral Graph Clustering](#)

*Carey E. Priebe, Johns Hopkins University*

2:00 PM

[Modeling Topics in Survey Interviewer Notes](#)

*Wendy Martinez, U.S. Bureau of Labor Statistics*

2:30 PM

[Eigen-Privy: Adjacency Spectral Embedding for Document Analysis](#)

*David Marchette, Naval Surface Warfare Center*



CS35 - Advances in Bayesian Analytics

Invited

Fri, May 18, 1:30 PM - 3:00 PM

Grand Ballroom E

*Organizer(s): Refik Soyer, George Washington University*

*Chair(s): Refik Soyer, George Washington University*

1:30 PM

[Deep Learning: A Bayesian Perspective](#)

*Vadim Sokolov, George Mason University*

2:00 PM

[Bayesian Analysis of Multivariate Non-Gaussian Time Series](#)

*Refik Soyer, George Washington University*

2:30 PM

[Likelihood, Confirmational Tenacity, and Mood Transitions in Bayesian Inference](#)

*Nozer D. Singpurwalla, City University of Hong Kong*



## CS36 - Data Visualization Platforms

Invited

Fri, May 18, 1:30 PM - 3:00 PM

Grand Ballroom F

*Organizer(s): Jim Harner, West Virginia University*

*Chair(s): Jim Harner, West Virginia University*

1:30 PM

[Using Shiny to interact with data](#)

*Winston Chang, Rstudio*

2:00 PM

[The Interactive Solution Path in JMP Pro: A Powerful Tool for Visualizing and Exploring Model Diagnostics](#)

*Chris Gotwalt, JMP*

2:30 PM

[RCloud - Collaborative Platform for Visualization and Data Analysis](#)

*Simon Urbanek, ATT Research*



## CS37 - Statistical Analytics for Data Science

Invited

Fri, May 18, 1:30 PM - 3:00 PM

Grand Ballroom G

*Organizer(s): Lynne Billard, University of Georgia*

*Chair(s): Seyed Yaser Samadi, Southern Illinois University Carbondale*

1:30 PM

[Time Series Analysis for Symbolic Interval-valued Data](#)

*Seyed Yaser Samadi, Southern Illinois University Carbondale*

2:00 PM

[Privacy Analytics via Aggregate Data: Trade-off between Statistical Efficiency and Privacy](#)

*Anand N. Vidyashankar, George Mason University*

2:30 PM

[Clustering Histogram-valued Data](#)

*Lynne Billard, University of Georgia*



## CS38 - Statistical Challenges in Large-Scale Data Mining

Invited

Fri, May 18, 1:30 PM - 3:00 PM



## Regency Ballroom A

*Organizer(s): Tian Zheng, Columbia University*

*Chair(s): Tian Zheng, Columbia University*

1:30 PM

[A Scalable Algorithm for Change-Points Computation in Large Graphical Models](#)

*Yves Atchade, University of Michigan*

2:00 PM

[Embedding Approaches for Mining Heterogeneous Information Networks](#)

*Yizhou Sun, UCLA*

2:30 PM

[Approximate Data Analytics](#)

*Christopher Jermaine, Rice University*



CS39 - Applications of Divide and Recombine to Big Data

Invited

Fri, May 18, 1:30 PM - 3:00 PM

Lake Fairfax A

*Organizer(s): William S. Cleveland, Purdue*

*Chair(s): Soren Harner, MuleSoft*

1:30 PM

[Divide & Recombine \(D&R\) with DeltaRho for Big Data Analysis](#)

*William S. Cleveland, Purdue*

2:00 PM

[DeltaRho for Deep Analysis of Precipitation and Cloud Observations to Advance the Understanding of Earth's Water Cycle](#)

*Wen-wen Tung, Earth, Atmospheric, and Planetary Sciences, Purdue*

2:30 PM

[Applications of Large-Scale Visualization Using Trelliscope](#)

*Ryan Hafen, Hafen Consulting LLC*



CS40 - Data Science Foundations

Contributed

Fri, May 18, 1:30 PM - 3:00 PM

Lake Fairfax B

*Chair(s): Snehalata Huzurbazar, West Virginia University*

1:30 PM

[A Grammar for Reproducible and Painless Extract-Transform-Load Operations on Medium Data](#)

*Ben Baumer, Smith College*

1:45 PM

[Perspectives on Deep Learning and Deep Reasoning](#)

*Rich Haney, Big Data2 Consulting*

2:00 PM

[Defining the AIM: An Abstraction for Improving Machine Learning Prediction](#)

*VICTORIA STODDEN, University of Illinois Urbana-Champaign*

2:15 PM

[Sensemaking and Five Problems with Big Data Science](#)

*Michael Latta, Coastal Carolina University - YTMBA Research & Consulting*

2:30 PM

[Painless Computing Models for Ambitious Data Science](#)

*Hatef Monajemi, Stanford University*

2:45 PM

[A Paradigm for Research in Data Science](#)

*Vardan Papyan, Stanford*

PS06 - Survey Data

E-Poster

Fri, May 18, 3:00 PM - 3:45 PM

Regency Ballroom B

1

[Constrained Optimization for Survey Weights](#)

*Matthew R Williams, Substance Abuse and Mental Health Services Administration*

2

[Performance Evaluation of Machine Learning Algorithms by K-Fold and Leave-One-Out Cross Validation for Classification of Survey Write-in Responses](#)

*Andrea Roberson, U.S. Census Bureau*

3

[Looking Inward: Quality Audits for Demographic Programs at the U.S. Census Bureau](#)

*Richard Levy, US Census Bureau*

4

[Some Dimension Reduction Strategies for the Analysis of Survey Data](#)

*Jiaying Weng, University of Kentucky*

5

[Suggestion of the Confidence Interval of the Cronbach Alpha in Application to Complex Survey Data](#)

*Jihnhee Yu, University at Buffalo*

6

[Secure Distributed Computational Processing for Industry Statistical Data](#)

*Cavan Paul Capps, U.S. Census Bureau*



CS41 - Big Data and Data Science in Government, Public Policy, and the Health Sciences  
Invited

Fri, May 18, 3:30 PM - 5:00 PM  
Grand Ballroom D

*Organizer(s): Nozer D. Singpurwalla, City University of Hong Kong; Inez Zwetsloot, City University of Hong Kong*  
*Chair(s): Inez Zwetsloot, City University of Hong Kong*

3:30 PM

[Building Resilient Communities: Harnessing the Power of Data](#)

*Sallie Keller, Biocomplexity Institute of Virginia Tech*

4:00 PM

[Data Foundation for Defense Acquisition: How the Department of Defense Manages and Uses Data to Support Management and Decision-making on the High-value Major Defense Acquisition Programs](#)

*Nancy Spruill, OUSD(AT&L)/ARA*

4:30 PM

[On the the Role of Higher Order Topological Properties in Functionality of Complex Networks](#)

*Yulia Gel, UT Dallas*



CS42 - Invitation to Statistical Analysis and Data Mining

Invited

Fri, May 18, 3:30 PM - 5:00 PM

Grand Ballroom E

*Organizer(s): Jia Li, Pennsylvania State University*  
*Chair(s): Lynne Billard, University of Georgia*

3:30 PM

[Fitting High-Dimensional Function-on-Scalar Regression Models via a Functional Augmented ADMM](#)

*Matthew Reimherr, Penn State University*

4:00 PM

[Flexible Supervised Learning Techniques for Block-missing Data](#)

*Yufeng Liu, University of North Carolina at Chapel Hill*

4:30 PM

[Phyloclustering: A Model-Based Approach for Identifying Microbial Populations](#)

*Wei-Chen Chen, pbdR Core Team*



CS43 - Dynamic Structural Proteomics: Simulation, Visualization, and Nonparametric Estimation

Invited

Fri, May 18, 3:30 PM - 5:00 PM

Grand Ballroom F

*Organizer(s): Juergen Symanzik, Utah State University*  
*Chair(s): Daniel B. Carr, George Mason University*

3:30 PM

[Biomolecules in Motion: Sample-based Models of Dynamics Elucidating Function and Mechanisms in the Healthy and Diseased Cell](#)

*Amarda Shehu, George Mason University*

4:00 PM

[Local PCA and Extraction of Filamentary Structures](#)

*Wanli Qiao, George Mason University*

4:30 PM

[An Approach to Visualizing Simulated Protein Folding Energy Landscapes as a Function of Four to Six Principal Components](#)

*Daniel B. Carr, George Mason University*



CS44 - Data Science Platforms II

Invited

Fri, May 18, 3:30 PM - 5:00 PM

Grand Ballroom G

*Organizer(s): Jim Harner, West Virginia University*  
*Chair(s): Jim Harner, West Virginia University*

3:30 PM

[The Unified Analytics Platform: Unifying Big Data Workloads in Apache Spark](#)

*Hossein Falaki, Databricks*

4:00 PM

[Using Microsoft ML Server and Spark for Distributed Computation of Massive Computational Experiments in Data Science and Statistical Inference](#)

*Ali Zaidi, Microsoft AI and Research*

4:30 PM

[The SAS® Platform: Where Point and Click Users and Coders of All Languages Collaborate Seamlessly](#)

*Carlos Pinheiro, SAS & Data Science Tech Institute, France*



CS45 - Statistical Machine Learning Applications in Surveys

Invited

Fri, May 18, 3:30 PM - 5:00 PM

Regency Ballroom A

*Organizer(s): Wendy Martinez, U.S. Bureau of Labor Statistics*  
*Chair(s): Wendy Martinez, U.S. Bureau of Labor Statistics*

3:30 PM

[Classification and Regression Trees and Forests for Imputing Data from Sample Surveys](#)

*MoonJung Cho, U.S. Bureau of Labor Statistics*

4:00 PM

[Model-Assisted Survey Estimation With Modern Prediction Techniques](#)

*Jean Opsomer, Colorado State University*

4:30 PM

[Calling All Stakeholders: Developing a Demographic Statistical Redesign Agenda](#)

*Richard Levy, US Census Bureau*



CS46 - Data Sciences Applications for Critical Health Issues I

Invited

Fri, May 18, 3:30 PM - 5:00 PM

Lake Fairfax A

*Organizer(s): William F. Wiecek, SUNY Buffalo State*

*Chair(s): Jonathan Lindner, Center for Health and Social Research at SUNY Buffalo State*

3:30 PM

[Alcohol Abstainers versus Drinkers: Changes in Health Outcomes after 20 Years](#)

*Kelly S Marczynski, SUNY Buffalo State*

4:00 PM

[Making Data Speak to User Needs: The Anchor Institution Dashboard](#)

*Alban Morina, Center for Health and Social Research at SUNY Buffalo State*

4:30 PM

[Conceptualization Issues in Analyzing and Communicating Collective Impact Data](#)

*Karl Wende, Center for Health & Social Research at Buffalo State*



CS47 - Time-to-Event Models

Contributed

Fri, May 18, 3:30 PM - 5:00 PM

Lake Fairfax B

*Chair(s): Rida Moustafa, Walmart*

3:30 PM

[A Moving 2D Time Series Models](#)

*Silvey Shamsi, Ball State University*

3:45 PM

[A Tool to Facilitate Creation of Multiple Time-Based Intervals per Subject](#)

*Cynthia Sue Crowson, Mayo Clinic*

4:00 PM

[An Efficient Generalized Least Squares Algorithm for Periodic Regression With Autoregressive Errors](#)

*Jaechoul Lee, Boise State University*

4:15 PM

[Comparison of Emotional States by Time Series Connectivity Analysis of Brain Activity Data](#)

*Rui Liu, Louisiana Tech University*

4:30 PM

[Floor Discussion](#)



CS48 - Distinguished Colleagues of Edward Wegman: Modeling and Data Science

Invited

Fri, May 18, 5:15 PM - 6:15 PM

Grand Ballroom D

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Yasmin H. Said, George Mason University*

5:15 PM

[The Revival of Statistical Ranking Methods in The High Technology and Big Data Era: Some Recent Developments](#)

*Michael G. Schimek, Medical University of Graz*

5:45 PM

[Communicating with Data Using Transparent Models](#)

*Roy E. Welsch, MIT*



CS49 - Data Analytics Supporting Homeland Security

Invited

Fri, May 18, 5:15 PM - 6:15 PM

Grand Ballroom E

*Organizer(s): Eddie Fuller, West Virginia University and Homeland Security*

*Chair(s): Eddie Fuller, West Virginia University and Homeland Security*

5:15 PM

[Vast & Varied - Big Data at DHS](#)

*Aaron Mannes, Homeland Security*

5:45 PM

[Using Data Analytics to Support Disaster Response During Harvey and Irma: Social Media, Weather and Other Data Sources](#)

*Eddie Fuller, West Virginia University and Homeland Security*



## CS50 - Data Science Platforms III

Invited

Fri, May 18, 5:15 PM - 6:15 PM

Grand Ballroom G

*Organizer(s): Jim Harner, West Virginia University*

*Chair(s): Soren Harner, MuleSoft*

5:15 PM

[Intelligent Application Networks with MuleSoft and TensorFlow](#)

*Soren Harner, MuleSoft*

5:45 PM

[An Introduction to the Watson Data Platform](#)

*Bernie Beekman, IBM*



## CS51 - Predictive Big Data Analytics

Invited

Fri, May 18, 5:15 PM - 6:15 PM

Regency Ballroom A

*Organizer(s): Jim Harner, West Virginia University*

*Chair(s): Jim Harner, West Virginia University*

5:15 PM

[Interpretable Machine Learning](#)

*Patrick Hall, H2O.ai*

5:45 PM

[Big Data with R](#)

*Edgar Ruiz, Rstudio*



## CS52 - Outcomes from the SAMSI Climate Program

Invited

Fri, May 18, 5:15 PM - 6:15 PM

Lake Fairfax A

*Organizer(s): David Banks, SAMSI and Duke University*

*Chair(s): David Banks, SAMSI and Duke University*

5:15 PM

[Modeling Large Spatial Data: an Application in Air Quality Modeling](#)

*Yawen Guan, SAMSI*

5:45 PM

[Inference on the Future State of the Climate Through Combining Multiple Interdependent Climate Model Outputs With Observations Using Bayesian Hierarchical Models](#)

*Huang Huang, SAMSI*



CS53 - Sports and Game Analytics

Contributed

Fri, May 18, 5:15 PM - 6:15 PM

Lake Fairfax B

*Chair(s): Rida Moustafa, Walmart*

5:15 PM

[Predict Video Game Wheel Design Game Strategy](#)

*Mason Chen, Stanford OHS*

5:30 PM

[Apply Multivariate Data Mining on Playing Strategic Video Game](#)

*Patrick Giuliano, Morrill Learning Center*

5:45 PM

[Baseball Pitching and Swing Contact Modeling](#)

*Andrew Chen, University of San Francisco*

6:00 PM

[Predict Basketball Team Winning Record](#)

*Mason Chen, Stanford OHS*



CS54 - Dynamic Data Visualization

Contributed

Fri, May 18, 5:15 PM - 6:15 PM

Grand Ballroom F

*Chair(s): Chris Gotwalt, JMP*

5:15 PM

[Dynamic Data Visualization: Bringing Data to Life](#)

*Neil W Polhemus, Statgraphics Technologies, Inc.*

5:30 PM

[Effective Story Telling with Dynamic Data Visualizations](#)

*Ruth M Hummel, JMP/SAS*

5:45 PM

[Exploratory Data Analysis for Predictive Analytics](#)

*Mia Stephens, JMP/SAS*



6:00 PM

[Discussant](#)

*Chris Gotwalt, JMP*

Saturday, May 19

Exhibits Open

SDSS Hours

Sat, May 19, 7:30 AM - 10:30 AM

Regency Ballroom Foyer

Registration

SDSS Hours

Sat, May 19, 7:30 AM - 12:00 PM

Registration



CS55 - New Directions in Rank Data Aggregation and Modeling

Invited

Sat, May 19, 8:30 AM - 10:00 AM

Grand Ballroom D

*Organizer(s): Michael G. Schimek, Medical University of Graz*

*Chair(s): William F. Wiecek, SUNY Buffalo State*

8:30 AM

[The Bayesian Mallows Model for Analysing Ranks and Preference Data: From Genomics to Recommendation Systems](#)

*Valeria Vitelli, University of Oslo*

9:00 AM

[Detecting and Interpreting Median Constrained Bucket Orders Within the Kemeny Axiomatic Framework](#)

*Antonio D'Ambrosio, University of Naples Federico II*

9:30 AM

[Discussant](#)

*Michael G. Schimek, Medical University of Graz*



CS56 - Data Science and Machine Learning in Naval Applications

Invited

Sat, May 19, 8:30 AM - 10:00 AM

## Grand Ballroom G

*Organizer(s): Jeffrey L. Solka, Naval Surface Warfare Center*

*Chair(s): Avory Bryant, Naval Surface Warfare Center*

8:30 AM

### [Using Found Data – A Cautionary Tale](#)

*David A. Johannsen, Naval Surface Warfare Center - Dahlgren*

9:00 AM

### [NLP-assisted Scientometric Horizon Scanning](#)

*Stuart Bingham, NSWCCD AM&DA, Code A43*

9:30 AM

### [Human Motion Analysis Using Deep Learning for Potential Threats](#)

*Alex Feild, Naval Surface Warfare Center*



## CS57 - Analyses Using Complex Models

Invited

Sat, May 19, 8:30 AM - 10:00 AM

Regency Ballroom A

*Organizer(s): Yasmin H. Said, George Mason University*

*Chair(s): Wendy Martinez, U.S. Bureau of Labor Statistics*

8:30 AM

### [Robust Multivariate Outlier Diagnostics in Chemometrics with Application to Spectrally Overlapping Drugs](#)

*Aylin Alin, Dokuz Eylul University*

9:00 AM

### [Statistical Learning in Big Data Analytics](#)

*S. Ejaz Ahmed, Brock University*

9:30 AM

### [Recovery of Ruin Probability and Value at Risk from the Scaled Laplace Transform Inversion](#)

*Adetokunbo Fadahunsi, West Virginia University*



## CS58 - Data science in Climate and Weather Research

Invited

Sat, May 19, 8:30 AM - 10:00 AM

Lake Fairfax A

*Organizer(s): Stephan R. Sain, Jupiter*

*Chair(s): Soren Harner, MuleSoft*

8:30 AM

[Large and Non-stationary Spatial Fields: Quantifying Uncertainty in Climate Models](#)

*Douglas Nychka, National Center for Atmospheric Research*

9:00 AM

[Regional Climate Model Assessment via Spatio-temporal Modeling](#)

*Peter Craigmile, The Ohio State University*

9:30 AM

[Statistical Downscaling and Uncertainty Quantification with Bayesian Deep Learning](#)

*Thomas Vandal, Northeastern University*



CS59 - Visualizing Complex Data

Contributed

Sat, May 19, 8:30 AM - 10:00 AM

Grand Ballroom F

*Chair(s): Jim Harner, West Virginia University*

8:30 AM

[Monitoring a survey with Google](#)

*Oscar Centeno Mora, University of Costa Rica*

8:45 AM

[Quantitative Evaluation of Manufacturing Visualization via Data Fusion](#)

*Xiaoyu Chen, Virginia Tech*

9:00 AM

[Mixed Type Distribution Plots](#)

*Christopher Weld, College of William & Mary*

9:15 AM

[Insights into Reshoring from Big Data Visualization of Social Media Posts](#)

*Megan Eileen Moore, North Carolina State University*

9:30 AM

[Discussant](#)

*Jim Harner, West Virginia University*



CS60 - Time-based Models

Contributed

Sat, May 19, 8:30 AM - 10:00 AM

Lake Fairfax B

*Chair(s): Suchismita Goswami, Computational Data Science, George Mason University*

8:30 AM

[Bankruptcy Prediction Using Selective Under-Sampling and Multiple-Year Data: A Study on North American Companies](#)

*Son Nguyen, Bryant University*

8:45 AM

[Looking Into Recurrent Event Data](#)

*Bommae Kim, University of Virginia Health System*

9:00 AM

[Artificial Neural Networks and Time Series Decomposition for the Flood Prediction in Mohawk Watershed, New York](#)

*Katerina Tsakiri, Rider University*

9:15 AM

[Causal Inference from Observational Time Series Data](#)

*Iris Tu, LinkedIn*

9:30 AM

[Detection of Excessive Activities in Time Series of Graphs Using Scan Statistics](#)

*Suchismita Goswami, Computational Data Science, George Mason University*

9:45 AM

[Floor Discussion](#)



CS61 - Data Sciences Applications for Critical Health Issues II

Invited

Sat, May 19, 10:30 AM - 12:00 PM

Grand Ballroom F

*Organizer(s): Kelly S. Marczyński, SUNY Buffalo State*

*Chair(s): Karl Wende, Center for Health & Social Research at Buffalo State*

10:30 AM

[Assessing the Impact of Weighted and Unweighted Alcohol Availability](#)

*Alan M Delmerico, Center for Health and Social Research at Buffalo State*

11:00 AM

[Survey of College Student Substance Abuse Problems: Role of Access and Norms](#)

*Jonathan Lindner, Center for Health and Social Research at SUNY Buffalo State*

11:30 AM

[Geographic Assessment of Adolescent Activity Space](#)

*William F. Wiecek, SUNY Buffalo State*



CS62 - Machine Learning for Complex Data

Contributed

Sat, May 19, 10:30 AM - 12:00 PM

Grand Ballroom D

*Chair(s): David Marchette, Naval Surface Warfare Center*

10:30 AM

[A Classification Tree for Functional Data](#)

*Jan Gertheiss, Clausthal University of Technology*

10:45 AM

[Optimal Estimation for Varying Coefficient Model with Longitudinal Data](#)

*Xiaowu Dai, University of Wisconsin Madison*

11:00 AM

[Regression Trees and Ensemble Methods for Multivariate Outcomes](#)

*Evan Lee Reynolds, University of Michigan*

11:15 AM

[XPCA: Interval-Censored Copula Principal Component Analysis for Discrete and Continuous Features](#)

*Clifford Anderson-Bergman, Sandia National Laboratories*

11:30 AM

[The Two-to-Infinity Norm and Singular Subspace Geometry With Applications to High-Dimensional Statistics](#)

*Joshua Cape, Johns Hopkins University*

11:45 AM

[Floor Discussion](#)

680123689  
790147900  
462701581  
462701581  
358631009  
680133689

CS63 - Data Science in Practice

Contributed

Sat, May 19, 10:30 AM - 12:00 PM

Grand Ballroom G

*Chair(s): Soren Harner, MuleSoft*

10:30 AM

[From Statistics to Data Science Startup: Transformation Within a Large Research Organization](#)

*Gayle S Bieler, RTI International*

10:45 AM

[Applied Techniques for Machine Learning with Limited Data](#)

*Andrew Hoblitzell, IUPUI; Andrew Hoblitzell, Purdue University*

11:00 AM

[Data Moves in Data Science Education](#)

*Tim Erickson, Epistemological Engineering*

11:15 AM

[Spatial Analysis of Crowdsourced Mobile Data](#)

*Arnab Chakraborty, North Carolina State University*

11:30 AM

[The SOBER Algorithm: How to Squeeze Out Huge but Sparse Data for Making Individual Predictions](#)

*Barbara Hildegard Wolf, GfK SE*

11:45 AM

[Floor Discussion](#)



CS64 - Bioinformatics  
Contributed  
Sat, May 19, 10:30 AM - 12:00 PM  
Regency Ballroom A

*Chair(s): Suchismita Goswami, Computational Data Science, George Mason University*

10:30 AM

[Method Selection and Graphical Network: Applications to Gene Expression Data](#)

*Demba Fofana, University of Texas Rio Grande Valley*

10:45 AM

[Scalable and Flexible Probabilistic PCA for Large-Scale Genetic Variation Data](#)

*Sriram Sankararaman, UCLA*

11:00 AM

[Type-I Error Rate of a One-Way ANOVA in the Case of a Large Number of Factors With Small Replications](#)

*Sharad Silwal, Jefferson College of Health Sciences*

11:15 AM

[Big Data Distributed System for Phenome and Genome Management and Analysis in a Large Health System](#)

*Wendy S.W. Wong, Inova Translational Medicine Institute*

11:30 AM

[Floor Discussion](#)



CS65 - Scientific and Financial Modeling  
Contributed  
Sat, May 19, 10:30 AM - 12:00 PM  
Lake Fairfax A

*Chair(s): Adetokunbo Fadahunsi, West Virginia University*

10:30 AM

[Apply Multivariate Statistics to study the Chocolate Science and Cardiovascular or Neurovascular Disease](#)

*Patrick Giuliano, Morrill Learning Center*

10:45 AM

[Probabilistic Particle-Filter Modeling of Shark Movement for Behavioral and Ecological Inference](#)

*Samuel Ackerman, Temple University*

11:00 AM

[Anomaly Detection in News Articles for Biosurveillance](#)

*Karl Pazdernik, Pacific Northwest National Laboratory*

11:15 AM

[Data Driven Portfolio Optimization Utilizing Machine Learning](#)

*Melinda Hsieh, Rider University*

11:30 AM

[Conditional Granger Causality Tests in Quantile Regression](#)

*Hong Cheng, Shanghai Lixin University of Accounting and Finance*

11:45 AM

[Floor Discussion](#)



CS66 - Business Analytics

Contributed

Sat, May 19, 10:30 AM - 12:00 PM

Lake Fairfax B

*Chair(s): Redouane Betrouni, George Mason University*

10:30 AM

[Modeling Emotions in Behavioral Big Data: Self-Selection, Impact Measures and Counterfactual Approach](#)

*Furio Camillo, University of Bologna*

10:45 AM

[Customer Perception Analysis using Statistical Modeling](#)

*Sridhar Ramaswamy, Caterpillar Inc*

11:00 AM

[Identifying and Utilizing Research Topics in Conference Abstracts](#)

*Stanislav Kolenikov, Abt Associates*

11:15 AM

[Forecasting Accuracy of Topic Modeling Techniques with Online Reviews: A Benchmark Study](#)

*Yuan Cheng, Cornell University*

11:30 AM

[Lookalike Audience Modeling with SVD and Cosine-Similarity](#)

*Sam Hawala, Resonate-Networks*

11:45 AM

[Predicting the Sale Price of Homes](#)

*Matea Milojkovic, Winthrop University*



CS67 - Feature Selection

Contributed

Sat, May 19, 1:15 PM - 2:45 PM

Grand Ballroom D

*Chair(s): Soren Harner, MuleSoft*

1:15 PM

[Statistical Testing for Feature Relevance: The HARVEST Algorithm](#)

*Herbert I Weisberg, Causalytics LLC*

1:30 PM

[Supervised Clustering via an Implicit Network for High Dimensional Data](#)

*Brandon Woosuk Park, George Mason University*

1:45 PM

[Variable Selection for the Recurrent Event Data with Broken Adaptive Ridge Regression](#)

*Dayu Sun, University of Missouri-Columbia*

2:00 PM

[Feature Selection in L0 Norm: A Viable Approach](#)

*Ana Maria Kenney, Pennsylvania State University*

2:15 PM

[Robust Surrogate Ridge Estimators for Linear Regression Model Based on an M-Estimator and MM-Estimator](#)

*Osama A Hussien, Alexandria University Egypt*

2:30 PM

[Floor Discussion](#)



CS68 - Data Science in Health

Contributed

Sat, May 19, 1:15 PM - 2:45 PM

Grand Ballroom G

*Chair(s): Kelly S Marczynski, SUNY Buffalo State*

1:15 PM

[Classifying Health Insurance Type from Survey Responses Using Enrollment Data](#)

*Joanne Pascale, US Census Bureau*

1:30 PM

[The Story of Goldilocks and Three Twitter APIs](#)

*Yoonsang Kim, NORC at the University of Chicago*

1:45 PM

[An Analysis of Crash-Safety Ratings and the True Assessment of Injuries by Vehicle](#)

*Cody Philips, Indiana University*

2:00 PM

[Association of Primary Tumor Site With Mortality in Patients Receiving Bevacizumab and Cetuximab for Metastatic Colorectal Cancer](#)

*Mayada Aljehani, Loma Linda University*

2:15 PM

[A Proposed Framework to Assess the Sensitivity of Network-Based Estimands to Non-Ignorable Non-Response, for Networks Ascertained With Non-Ignorable Sampling](#)

*Kenneth J Wilkins, National Institutes of Health, National Institute of Diabetes & Digestive & Kidney Diseases*

2:30 PM

[Floor Discussion](#)





CS69 - Image and High-Dimensional Processing  
Contributed  
Sat, May 19, 1:15 PM - 2:45 PM  
Regency Ballroom A

*Chair(s): Adetokunbo Fadahunsi, West Virginia University*

1:15 PM

[Analysis of Diagnostic Tests in the CTC Images for Detecting Colon Polyps](#)

*Krishna K Saha, Central CT State University*

1:30 PM

[Sequential Multi-Aspect Monitoring Multivariate and High-Dimensional Data](#)

*Amitava Mukherjee, XLRI - Xavier School of Management*

1:45 PM

[Robust Analysis of High Dimensional Data](#)

*Quefeng Li, UNC Chapel Hill*

2:00 PM

[Anisotropic Functional Laplace Deconvolution](#)

*RASIKA RAJAPAKSHAGE, University of Central Florida*

2:15 PM

[Bayesian Variable Selection Using Spike and Slab Prior With Application to High Dimensional EEG Data by Local Modeling](#)

*Shariq Mohammed, University of Connecticut*

2:30 PM

[Floor Discussion](#)



CS70 - Public Health Applications  
Contributed  
Sat, May 19, 1:15 PM - 2:45 PM  
Lake Fairfax A

*Chair(s): Redouane Betrouni, George Mason University*

1:15 PM

[A Comprehensive Analysis of Trends and Determinants of HIV/AIDS Knowledge Among the Bangladeshi Women Based on Bangladesh Demographic and Health Surveys, 2007–2014](#)

*Md. Tuhin Sheikh, Department of Statistics, University of Connecticut, Storrs, CT.*

1:30 PM

[Markov Process Multistate Modeling of Large Data](#)

*Marepalli B Rao, University of Cincinnati*

1:45 PM

[Approaches to Investigating Multimorbidity](#)

*Cynthia Sue Crowson, Mayo Clinic*

2:00 PM

[A New Framework for Re-identification Risk Estimation in Complex Healthcare Data](#)

*Lei Li, George Mason University*

2:15 PM

[Application of Algorithms With Serial Hepatitis C RNA Tests to Predict Treatment and Sustained Virologic Response Among Patients Infected With Chronic Hepatitis C.](#)

*Ademola B Osinubi, Centers for Disease Control and Prevention*

2:30 PM

[Floor Discussion](#)

GS04 - Closing Remarks

General Session

Sat, May 19, 3:00 PM - 3:30 PM

Grand Ballroom D

*Organizer(s): Jim Harner, West Virginia University; Yasmin H. Said, George Mason University*