

Printable Program
Printed on: 08/19/2025

Wednesday, June 2

SDSS Virtual Expo
SDSS Hours
Wed, Jun 2, 11:30 AM - 6:30 PM

GS01 - Welcome and Plenary Panel: Equitable and Inclusive Data and Technology
General Session
Wed, Jun 2, 11:45 AM - 1:00 PM

Chair(s): Wendy Martinez, Bureau of Labor Statistics

[Equitable and Inclusive Data and Technology](#)

*Safiya Noble, University of California, Los Angeles; Stephen T. Ziliak, Roosevelt University;
Jamelle Watson-Daniels, Data for Black Lives*



CS01 - Data Science Shaping the Financial World
Refereed
Wed, Jun 2, 1:10 PM - 2:45 PM

Chair(s): Kali Prasun Chowdhury, University of California, Irvine

1:15 PM

[BERT as a Filter to Detect Pharmaceutical Innovations in News Articles](#)

Martha Czernuszenko, The University of Virginia

1:45 PM

[Dissecting the 2015 Chinese Stock Market Crash](#)

Min Shu, University of Wisconsin-Stout

2:15 PM

[A Hierarchical Bayesian Approach to Detecting Structural Changes in Bank Liquidity Premia](#)

Padma Ranjini Sharma, Federal Reserve Bank of Kansas City



CS02 - Bayesian Approaches
Refereed
Wed, Jun 2, 1:10 PM - 2:45 PM

Chair(s): Claire Bowen, Urban Institute

1:15 PM

[Parameter Estimation for Ising Model with Variational Bayes](#)

Minwoo Kim, Michigan State University

1:45 PM

[Bayesian Heteroskedasticity-Robust Regression](#)

Gabriel Durbin Lewis, University of Massachusetts Amherst

2:15 PM

[A Scalable Bayesian Hierarchical Modeling Approach for Large Spatio-Temporal Count Data](#)

Aritz Adin, Public University of Navarre



CS03 - Shaping Decisions with Classification and Clustering

Refereed

Wed, Jun 2, 1:10 PM - 2:45 PM

Chair(s): Thomas Chen, Academy for Mathematics, Science, and Engineering

1:15 PM

[Unbiased Estimations Based on Binary Classifiers: A Maximum Likelihood Approach](#)

Marco J.H. Puts, Statistics Netherlands

1:45 PM

[Spectral Clustering of Mixed-Type Data](#)

Felix Mbuga, San Jose State University

2:15 PM

[Cluster Analysis in a Multi-Block Setting](#)

Fabien Llobell, Addinsoft, XLSTAT

CS04 - Late Breaking: Recent Trends in High-Dimensional Statistics

Special Session

Wed, Jun 2, 1:10 PM - 2:45 PM

Chair(s): Wendy Martinez, Bureau of Labor Statistics

1:15 PM

[On Graphical Models and Convex Geometry](#)

Haim Bar, University of Connecticut

1:45 PM

[High-Dimensional Latent Factor Modeling with Missing Observations](#)

Ruoxuan Xiong, Stanford University

2:15 PM

[James-Stein Estimators for Sample Eigenvectors](#)

Alex Shkolnik, University of California, Santa Barbara



CS05 - Assessing the Impact of COVID-19 Across Domains

Lightning

Wed, Jun 2, 1:10 PM - 2:45 PM

Chair(s): Brennan Bean, Utah State University

1:15 PM

[Monitoring Demographics of Enrolling Participants in COVID-19 Phase 3 Vaccine Studies using Tableau](#)

Rohit Banerjee, Fred Hutchinson Cancer Research Center

1:20 PM

[WITHDRAWN Covid-19 Trends and Analysis in the New England Region](#)

1:25 PM

[Exploring the Effects of Various U.S. Policies since the COVID-19 Outbreak using SIR-based Models](#)

David Han, The University of Texas at San Antonio

1:30 PM

[Aggregating Statistical Models and Human Judgment to forecast COVID-19](#)

Damon Luk, Lehigh University

1:35 PM

[Forecasting COVID-19 Hospital Census: A Multivariate Time-series Model Based on Local Infection Incidence](#)

Hieu Minh Nguyen, Center for Outcomes Research and Evaluation, Atrium Health

1:40 PM

[Exploring Business Operations through Network Visualization](#)

Joshua Logan Colburn, SciLine, AAAS

1:45 PM

[The Impact of COVID-19 on Elected Official Rhetoric and Framing of Healthcare Issues](#)

Catherine C. Pollack, Dartmouth College

1:50 PM

[The Interplay of Demographic Variables and Social Distancing Scores in the Deep Prediction of US Covid-19 Cases](#)

Francesca Tang, Princeton University

1:55 PM

[Age and Gender Perspective of COVID-19 Mortality in San Diego County](#)

Sarah Jing Dong, Westview High School

2:00 PM

[What Most Affects the Probability of Receiving Public Assistance? Examining the Effect of Family Background and Educational Attainment on Receiving Public Assistance with Multivariate Regression](#)

Patricia Vargas, Oregon State University

2:05 PM

[Measuring the progression of COVID-19 on social media](#)

Marco J.H. Puts, Statistics Netherlands

2:10 PM

[WITHDRAWN COVID-19 and Financial News: Using Text Analytics to Assess Bank Networks During Crisis](#)

2:15 PM

[WITHDRAWN 2020 United States Presidential Election Prediction Model](#)

2:20 PM

[Network Group Testing](#)

Paolo Bertolotti, MIT

2:25 PM

[Machine Learning Methods for Online Panel-based Surveys](#)

Yulei He, CDC

2:30 PM

[Network Structure of Cross-Currency Swap Market During the COVID-19 Crisis](#)

Kazuaki Washimi, Bank of Japan

2:35 PM

[WITHDRAWN An Integer-Valued GARCH Process to Model Cases of COVID-19](#)

SC1 - Data Visualization with R, Part 1

Short Course

Wed, Jun 2, 3:00 PM - 6:30 PM

Instructor(s): Aaron Robert Williams, Urban Institute

Data visualization plays a crucial role in the data science and statistics workflows. It is fundamental to everything from exploratory data analysis to communicating results. Data scientists and statisticians can better understand data and more effectively communicate their work by understanding how to better visualize their data. Too often, however, visualization is an afterthought.

In this course, attendees will learn the core principles of data visualization how we perceive visual information; the layered grammar of graphics; and best practices for creating effective visualizations. To put these principles to work, attendees will learn practical skills for R programming that improve the quality of their work and teach them to program away the mundane. The course will focus on the popular R package ggplot2 and the reproducible research framework R Markdown. All R instruction will begin with a clear motivation, followed by an explanation of the approach and code and ending with hands-on examples.

SC2 - Deep Learning in Statistics, Part 1

Short Course

Wed, Jun 2, 3:00 PM - 6:30 PM

Instructor(s): Edgar Dobriban, University of Pennsylvania; Annie Qu, University of California, Irvine; Xiao Wang, Purdue University

This short course is for those who are new to data science and interested in understanding the cutting-edge machine learning and deep learning models. It is for those who want to become familiar with the core concepts behind these learning algorithms and their successful applications and who want to start thinking about how machine learning and deep learning might be useful in their research, business, or career development. The course will provide a comprehensive overview of statistical machine learning and deep learning methods. Topics include classical methods and

modern techniques, including basic machine learning tools, supervised and unsupervised learning, deep neural network, computational algorithms and software of deep learning, and various applications in deep learning.

SC3 - Artificial Intelligence, Machine Learning, and Precision Medicine
Short Course
Wed, Jun 2, 3:00 PM - 6:30 PM

Instructor(s): Haoda Fu, Eli Lilly

In this half-day short course, I will provide an overview of statistical machine learning and artificial intelligence techniques, with applications to precision medicine, particularly deriving optimal individualized treatment strategies for personalized medicine. We will cover both treatment selection and treatment transition. The treatment selection framework is based on outcome-weighted classification. We will discuss logistic regression, support vector machine (SVM), λ -learning, robust SVM, and angle-based classifiers for multi-category learning. I will show how to modify these classification methods into outcome-weighted learning algorithms for personalized medicine. The second part of this course will cover treatment transition. I will provide an introduction to reinforcement learning techniques. Algorithms—including dynamic programming for Markov decision process, temporal difference learning, SARSA, Q-Learning algorithms, and actor-critic methods—will be covered. We will discuss how to use these methods for developing optimal treatment transition strategies. The techniques discussed will be demonstrated in R. This course is intended for graduate students who have some knowledge of statistics and want to be introduced to statistical machine learning, or practitioners who would like to apply statistical machine learning techniques to their problems on personalized medicine and other biomedical applications.

Thursday, June 3

SDSS Virtual Expo
SDSS Hours
Thu, Jun 3, 9:30 AM - 6:30 PM



CS06 - Shaping Human Health with Data
Refereed
Thu, Jun 3, 10:00 AM - 11:35 AM

Chair(s): Julia Buffinton, LMI

10:05 AM

[Prescription Opioid Epidemic: Current Trends, Analysis, and Interpretation](#)

Thomas Bryan, BGD

10:35 AM

[Predicting Adverse Drug Reactions \(ADR\) Using Physiological Time Series Data Obtained in the Intensive Care Unit \(ICU\): A Case Study](#)

Jason Lee, Johns Hopkins University Applied Physics Laboratory

11:05 AM

[SVM-Based Models for Pill Shape Classification](#)

William Franz Lamberti, George Mason University



CS07 - Estimation Techniques

Refereed

Thu, Jun 3, 10:00 AM - 11:35 AM

Chair(s): Jean Opsomer, Westat

10:05 AM

[A New Sufficient Dimension Reduction Predictive Model Using Maximum Entropy Covariance Estimator with Information Complexity](#)

Kabir Opeyemi Olorede, Kwara State University

10:35 AM

[Quantile Shrinkage Covariance Estimation](#)

qiyu wang, Hong Kong Polytechnic Univ

11:05 AM

[Nonparametric Application of Functional Analysis of Generalized Linear Models Under Nonlinear Constraints](#)

Kali Prasun Chowdhury, University of California, Irvine



CS08 - Network Analysis

Refereed

Thu, Jun 3, 10:00 AM - 11:35 AM

Chair(s): Katie Anne Bakewell, NLP Logix

10:05 AM

[Estimation of the Mean Function of Functional Data via Deep Neural Networks](#)

SHUOYANG WANG, Auburn University

10:35 AM

[Interpretation of Radiological Imaging Features Using Generative Adversarial Networks](#)

Kyle Andrew Hasenstab, San Diego State University

11:05 AM

[Multinomial Tensor Regression with Application to Whole-Brain Structural MRI Analysis](#)

Fang Yang, University of Cincinnati



CS09 - Online/Hybrid Teaching in Statistics and Data Science

Special Session

Thu, Jun 3, 10:00 AM - 11:35 AM

Chair(s): Laura Le, University of Minnesota

10:05 AM

[Lessons for the Future from a Year of COVID Teaching](#)

Steven Foti, University of Florida; Adam Loy, Carleton; Douglas Whitaker, Mount Saint Vincent University; Laura Ziegler, Iowa State University



CS10 - Classification and Simulation: Methods, Analyses, and Applications

Lightning

Thu, Jun 3, 10:00 AM - 11:35 AM

Chair(s): Stanislav Kolenikov, Abt Associates

10:05 AM

[Auto-classification of occupational data](#)

Ning Chong, Ministry of Manpower

10:10 AM

[Identifying different types of companies via their website text](#)

Piet J. Daas, Eindhoven University of Technology

10:15 AM

[Identifying the Tone of FOMC Statements Using a Natural Language Processing Tool](#)

Taeyoung Doh, Federal Reserve Bank of Kansas City

10:20 AM

[Using Data Visualization to Tell the Complex Story of Language Use Trends in United States](#)

Heather Hisako Kitada Smalley, Willamette University

10:25 AM

[How SciLine Solved its Multi-Label Classification Problem](#)

Joshua Logan Colburn, SciLine, AAAS

10:30 AM

[An Extension of DEIM for Class Identification](#)

Emily Hendryx, University of Central Oklahoma

10:35 AM

[Character-Level Representation Model for Domain-Specific Short Text](#)

Xinming Li, Global Tech, Walmart

10:40 AM

[An Algorithm for Discrete Logistic Classification for Sparse Tables](#)

Yves Thibaudeau, U.S. Census Bureau

10:45 AM

[Classification of Longitudinal Data with Irregularly Spaced Intervals: Mixture-Based Mixed Effects Models Versus Post-Hoc Mixture Models of the Best Linear Unbiased Predictors \(BLUP\) from Linear Mixed](#)

Md Jobayer Hossain, Nemours Children Hospital System, A. I. Dupont Hospital for Children

10:50 AM

[Robust Meta-Analysis for Large-Scale Genomic Experiments based on an Empirical Approach](#)

Sinjini Sikdar, Old Dominion University

10:55 AM

[Disease Associated Network Detection in Multi-Omic Single-Cell Experiments](#)

Lorin Towle-Miller, University at Buffalo

11:00 AM

[An empirical Bayes approach to estimating dynamic models of co-regulated gene expression](#)

Sara Venkatraman, Cornell University

11:05 AM

[Pathway and Gene Selection with Guided Regularized Random Forests](#)

Tyler Cook, University of Central Oklahoma

11:10 AM

[WITHDRAWN Prediction Interval of Air Pollutants Concentration by Nonparametric Regression Analysis](#)

11:15 AM

[Bayesian wavelet-packet historical functional linear models](#)

Mark J Meyer, Georgetown University

11:20 AM

[Using Simulation-Based Inference to mitigate instrumental biases in X-ray telescopes](#)

Daniela Huppenkothen, SRON Netherlands Institute for Space Research

11:25 AM

[Regional and Sectoral Structures and Their Dynamics of the Chinese Economy: A Network Perspective from Multi-Regional Input-Output Tables](#)

Jun Yan, University of Connecticut

11:30 AM

[Connecting the Popularity Adjusted Block Model to the Generalized Random Dot Product Graph](#)

John Koo, Indiana University



CS11 - Administrative Data Analysis Shaping Decisions

Refereed

Thu, Jun 3, 1:10 PM - 2:45 PM

Chair(s): Mian Shams Adnan, Bowling Green State University

1:15 PM

[Learning About Homelessness Using Linked Survey and Administrative Data](#)

Angela Jean Wyse, University of Chicago

1:45 PM

[Equitable Prediction of Suicide from Administrative Patient Records](#)

Majerle Reeves, University of California, Merced

2:15 PM

[Disambiguating Patent Inventors, Assignees, and Their Locations in PatentsView](#)

Christina Jones, American Institutes for Research



CS12 - Software and Technology Shaping Data Science

Refereed

Thu, Jun 3, 1:10 PM - 2:45 PM

Chair(s): Tommy Jones, Data Community DC

1:15 PM

[Synthetic Data Generation with Tidysynthesis](#)

Aaron Robert Williams, Urban Institute

1:45 PM

[Estimation and Clustering in the Sparse Popularity Adjusted Blockmodel](#)

Ramchandra Rimal, Middle Tennessee State University

2:15 PM

[The Importance of Good Coding Practices for Data Scientists](#)

Randall Pruim, Calvin University



CS13 - Visual Analytics

Refereed

Thu, Jun 3, 1:10 PM - 2:45 PM

Chair(s): Brennan Bean, Utah State University

1:15 PM

[VitalVis: Visual Analysis of Multivariate Time Series Data for Healthcare](#)

Chad A. Steed, Oak Ridge National Laboratory

1:45 PM

[Validating Visual Inference Methods by Use of Deep Learning](#)

Anne Helby Helby Petersen, University of Copenhagen

2:15 PM

[Using Geographic Information Systems \(GIS\) and Spatial Statistics to Support the Education to Workforce Pipeline and Address Inequities](#)

Caitlin Deal, American Institutes for Research



CS14 - Data-Driven Healthcare

Lightning

Thu, Jun 3, 1:10 PM - 2:45 PM

Chair(s): Wendy Martinez, Bureau of Labor Statistics

1:15 PM

[Long-Term and High-Exposure Effects of NB-UVB Phototherapy in Relation to Increased Risks of Skin Cancers: A 30-Year Cancer Registry Linkage Study](#)

Khaled Bedair, Faculty of Commerce, Tanta University

1:20 PM

[CNVScope: Visually Exploring Copy Number Aberrations in Cancer Genomes](#)

James L T Dalgleish, National Cancer Institute, Center for Cancer Research, Genetics Branch

1:25 PM

[Temporal Prediction of Future Disease States using High-dimensional Covariates](#)

Sandipan Dutta, Old Dominion University

1:30 PM

[Creation of Breast Cancer Subtypes Through a Consensus-Based Network Approach](#)

Christina Horr, University of Notre Dame

1:35 PM

[Did Increasing Continuity of Care Protect Patients with Chronic Disease from Emergency and Hospitalization Readmission? A Cohort Spatial-Temporal Study in Mississippi](#)

Phi Le, University of Mississippi Medical Center

1:40 PM

[NCHS Data Linkage Program: Leveraging the Nation's Health Data for Evidence-Based Decision-Making](#)

Lisa B. Mirel, CDC/NCHS

1:45 PM

[Real-Time Client Attrition Prediction in the Nurse Family Partnership Home Visiting Program](#)

Kaushik Mohan, Two Sigma

1:50 PM

[Empirical Calibration of a Simulation Model of Opioid Use Disorder](#)

Anusha Madushani Rajapaksha Wasala Mudiyansele, Boston Medical Center

1:55 PM

[Random Survival Forests for Dynamic Prediction of a Time-to-event outcome using a Longitudinal Biomarker](#)

Krithika Suresh, University of Colorado

2:00 PM

[County-level Low Birth Weight Rates and Associated Contextual Factors in the United States, 2011-2016](#)

Pallavi Dwivedi, University of Maryland College Park

2:05 PM

[A Population-Based Study of Associations Between Attainment of Incentivized Primary Care Indicators and Emergency Hospital Admissions Among Those with Type 2 Diabetes in England](#)

Laura H Gunn, University of North Carolina at Charlotte & Imperial College London

2:10 PM

[Important factors to predict Anemia during the treatment of Malaria in HIV-infected population](#)

Yein Jeon, Georgetown University

2:15 PM

[Identification of latent relationships between disability rates and socio-geographic variables in veterans utilizing Machine Learning methods](#)

Gina McKernan, University of Pittsburgh

2:20 PM

[Survival Analysis Based on Statistical Modeling Versus Cox Proportional Hazard Model of Multiple Myeloma Cancer Patients](#)

Lohuwa Mamudu, University of South Florida

2:25 PM

[Sequential Pattern Mining of Electronic Health Record for Early Diagnosis of Amyotrophic Lateral Sclerosis](#)

Lily Sun, Stanford OHS

2:30 PM

[Trinary \(+/0/-\) Categorization for Tracing Step-Based Shifts over Time and Identifying Hot Spots in Big Data](#)

Turkan Kumbaraci Gardenier, Teka Trends, Inc.

2:35 PM

[Bayesian Estimation of Program-Specific Impacts in the HPOG Program](#)

Stanislav Kolenikov, Abt Associates

2:40 PM

[Image clustering of brain tumor patients using 3D convolutional auto-encoder](#)

Seyed Mohammad Hadi Hosseini, St. Jude Children's Research Hospital



CS15 - Addressing Big Data Challenges: Topics in Deep Learning and Model Monitoring
Lightning

Thu, Jun 3, 1:10 PM - 2:45 PM

Chair(s): SHUOYANG WANG, Auburn University

1:15 PM

[WITHDRAWN Visual Similarity in Ranking E-Commerce Listings](#)

1:20 PM

[Automated Active Monitoring of Production Machine Learning Models](#)

Katie Anne Bakewell, NLP Logix

1:25 PM

[WITHDRAWN A Novel Machine Learning Approach for Humanitarian Relief Assistance After Natural Disasters](#)

1:30 PM

[A Brief Review of Quantum Computation, Quantum Algorithms, and Impact on Practice of Data Science](#)

David Han, The University of Texas at San Antonio

1:35 PM

[Score-Based Change Detection for Gradient-Based Learning Machines](#)

Lang Liu, University of Washington

1:40 PM

[Tolstoy Targets: An efficient niche graph.](#)

Dennis Sweitzer, YPrime

1:45 PM

[Monitoring of sunspot number observations based on neural networks](#)

Sophie Mathieu, UCLouvain

1:50 PM

[Utilizing stability criteria in choosing feature selection methods yields reproducible results in](#)

[microbiome data](#)

Lingjing Jiang, Johnson & Johnson

1:55 PM

[Self-Supervised Learning for Robust Image Classification](#)

Ladyna Wittscher, Friedrich-Schiller-Universität Jena

2:00 PM

[Bayesian forward modeling of high-resolution radio interferometric gravitational lens observations](#)

Devon Powell, Max Planck Institute for Astrophysics

2:05 PM

[Modeling Implicit Feedback in Visual Recommendations for E-Commerce](#)

Julia Zhou, Etsy, Inc

2:10 PM

[Contextual Matching via Graph Representation Learning with Side Information](#)

Chris Xu, Etsy

2:15 PM

[Loss convergence in a causal Bayesian neural network of retail firm performance](#)

F. Trevor Rogers, University of Hawaii, Manoa

2:20 PM

[WITHDRAWN Statistically Structured Computations vs. Perceptron Computations: New Opportunities](#)

SC1 - Data Visualization with R, Part 2

Short Course

Thu, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Aaron Robert Williams, Urban Institute

Data visualization plays a crucial role in the data science and statistics workflows. It is fundamental to everything from exploratory data analysis to communicating results. Data scientists and statisticians can better understand data and more effectively communicate their work by understanding how to better visualize their data. Too often, however, visualization is an afterthought.

In this course, attendees will learn the core principles of data visualization how we perceive visual information; the layered grammar of graphics; and best practices for creating effective visualizations. To put these principles to work, attendees will learn practical skills for R programming that improve the quality of their work and teach them to program away the mundane. The course will focus on the popular R package ggplot2 and the reproducible research framework R Markdown. All R instruction will begin with a clear motivation, followed by an explanation of the approach and code and ending with hands-on examples.

SC2 - Deep Learning in Statistics, Part 2

Short Course

Thu, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Edgar Dobriban, University of Pennsylvania; Annie Qu, University of California, Irvine; Xiao Wang, Purdue University

This short course is for those who are new to data science and interested in understanding the cutting-edge machine learning and deep learning models. It is for those who want to become familiar with the core concepts behind these learning algorithms and their successful applications and who want to start thinking about how machine learning and deep learning might be useful in their research, business, or career development. The course will provide a comprehensive overview of statistical machine learning and deep learning methods. Topics include classical methods and modern techniques, including basic machine learning tools, supervised and unsupervised learning, deep neural network, computational algorithms and software of deep learning, and various applications in deep learning.

SC4 - Data Quality for Data Science and Statistics: A Survey and Practical Application
Short Course
Thu, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Henry Li, Bigeye

Data science and statistics become more important in society every year—as a prime example, consider the sudden influx of public interest in COVID-19 tracking projects such as the tracker from 1Point3Acres. From published research that guides policy to the online predictive systems that set prices and control what we read, high-quality and reliable input data is a necessary (but not sufficient!) condition for quality outcomes.

This half-day course will cover the impact of data quality issues on data science and statistics work, taxonomies of data quality issues that can occur, a survey of current techniques and tools for issue identification, and how to start including data quality techniques in one's data science work process.

Friday, June 4

SDSS Virtual Expo
SDSS Hours
Fri, Jun 4, 9:30 AM - 6:30 PM

GS02 - Plenary Panel: Impacts of COVID-19
General Session
Fri, Jun 4, 10:00 AM - 11:10 AM

Chair(s): Donna LaLonde, American Statistical Association

[Impacts of COVID-19](#)

Lily Wang, Iowa State University; Sara Del Valle, Los Alamos National Laboratory; Juan Lavista Ferres, Microsoft



CS16 - Data Science Shaping Innovative Applications

Refereed

Fri, Jun 4, 11:25 AM - 1:00 PM

Chair(s): Jessica Moore, LMI

11:30 AM

[Community Detection in Open Source Software Collaboration Networks](#)

Behnaz Moradijamei, University of Virginia

12:00 PM

[Improving Election Predictions: A Statistical Investigation of Crucial Swing State Behavior in the Past](#)

Mason Chen, Stanford OHS



CS17 - Simulation-Based Statistics

Refereed

Fri, Jun 4, 11:25 AM - 1:00 PM

Chair(s): Sinjini Sikdar, Old Dominion University

11:30 AM

[Local Two-Sample Testing by Random-Walk Distributions](#)

Boris Landa, Yale University

12:00 PM

[K-Fold Cross-Validation for Complex Sample Surveys](#)

Jerzy Wieczorek, Colby College

12:30 PM

[The Impact of Bias Correction on Bootstrap Confidence Intervals for the First Order Bifurcating Autoregressive Model](#)

Tamer Elbayoumi, North Carolina A&T State University



CS18 - Advancements in Learning

Refereed

Fri, Jun 4, 11:25 AM - 1:00 PM

Chair(s): Heather Hisako Kitada Smalley, Willamette University

11:30 AM

[Adversarially Robust Subspace Learning](#)

Ruizhi Zhang, University of Nebraska-Lincoln

12:00 PM

[A Geometry-Driven Longitudinal Topic Model](#)

Yu Wang, University of Michigan

CS19 - Memorial Session for Jim Harner

Special Session

Fri, Jun 4, 11:25 AM - 1:00 PM

Chair(s): Tim Hesterberg, Google

11:30 AM

[How Usability and Reproducibility in Software Improves Teaching and Research](#)

Soren Harner, RC2AI

11:55 AM

[A Predictive Internet-Based Model for COVID-19 Hospitalization Census](#)

Philip Turk, Atrium Health

12:20 PM

[Leading with Data Science: A Key to West Virginia's Pandemic Response](#)

Brad Price, West Virginia University



CS20 - Bio-Data Science

Refereed

Fri, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Turkan Kumbaraci Gardenier, Teka Trends, Inc.

1:25 PM

[Examining Leading Indicators for COVID-19 Incidence Case Growth](#)

Vishnu Shankar, Stanford University

1:55 PM

[Propagating Uncertain Functional Inputs Through Neutronics Simulations](#)

Scott Vander Wiel, Los Alamos National Laboratory



CS21 - What Do I Use Computational Statistics For?

Refereed

Fri, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Donna LaLonde, American Statistical Association

1:25 PM

[A Statistical Analysis to Determine the Distribution and Pattern for an Insurance Claim Data](#)

Sikiru Ademola Adewale, Virginia Tech

1:55 PM

['To Book or Not to Book': Exploring the Determinants of Successful Airbnb Bookings Across Multiple Geographies](#)

Pooja Sengupta, International Management Institute, Kolkata

2:25 PM

[Sparse Bayesian Predictive Modeling of Tumor Response from Radiomic Data](#)

Shirin Golchi, McGill University



CS22 - Shaping the Next Generation

Refereed

Fri, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Claire Bowen, Urban Institute

1:25 PM

[Examples from Two Undergraduate Bayesian Courses: Tools, Resources, and Suggestions for Teaching](#)

Jingchen Hu, Vassar College

1:55 PM

[Computing Competencies for Undergraduate Data Science Curricula: Summary of Report and Next Steps](#)

Karl R. B. Schmitt, Trinity Christian College

2:25 PM

[Developing a High-School Data Science Course and Curriculum](#)

Daniel O'Loughlin, St. Paul Academy and Summit School



CS23 - Data Science Education and Applications

Lightning

Fri, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Lisa Kay, Eastern Kentucky University

1:25 PM

[Novel Statistical Analysis in the Context of a Comprehensive Needs Assessment for Secondary STEM Recruitment](#)

Norou Diawara, Old Dominion University

1:30 PM

[Who teaches data science concepts? - Results from course catalog mining with machine learning](#)

Sasha Lioutikova, Yale University

1:40 PM

[Using Data from my Backyard Garden to Teach Introductory Data Science](#)

Lisa Lendway, Macalester College

1:45 PM

[Scaffolding and Mentoring Student Data Science Projects in the Liberal Arts Setting](#)

Heather Hisako Kitada Smalley, Willamette University

1:50 PM

[Reproducible and collaborative data-science with the RENKU platform](#)

Oksana Riba, EPFL Swiss Data Science Center

1:55 PM

[Evaluating the overlap weighting method for balancing covariates using real and simulated claims data](#)

Wen Wan, The University of Chicago

2:00 PM

[Reconstruction of age distributions from differentially private population data](#)

Ethan Sharygin, Portland State University

2:05 PM

[Predicting Type of Work-Related Fatal Accident Based on Knowledge Graph and Machine-Learning Methods](#)

Feiyu E, Southern University of Science and Technology

2:10 PM

[Tornado Property Loss Scale: Up to \\$8 Billion by 2025. Classification, Dependence, and Prediction of Tornado Events in the U.S.](#)

Thilini Vasana Mahanama, Texas Tech University

2:15 PM

[Tax Auditing use of Statistical Sampling Design Methods](#)

Zachary Rhyne, Ryan LLC

2:20 PM

[The Effect of the Differential Privacy Disclosure Avoidance System Proposed by the Census Bureau on 2020 Census Products: Four Case Studies of Census Blocks in Alaska.](#)

David A. Swanson, UC Riverside and CSDE, U of Washington

2:25 PM

[Analysis of 2020 Global Stock Market Crash](#)

Min Shu, University of Wisconsin-Stout

2:30 PM

[Using Timescale Variation Measures to Measure Climate Change](#)

Yawen Ding, Cornell University



CS24 - New Models and Methods

Lightning

Fri, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Md Jobayer Hossain, Nemours Children Hospital System, A. I. Dupont Hospital for Children

1:25 PM

[Likert-type scale variables analysed through CUB models: a review](#)

Nicolò Biasetton, Università degli Studi di Padova

1:30 PM

[Machine Learning Assisted Complex Survey Weights](#)

Stanislav Kolenikov, Abt Associates

1:35 PM

[An Efficient Variance Estimator for Cross-Validation under Partition-Sampling](#)

Xizhen Cai, Williams College

1:40 PM

[Methods for Detecting Numeracy's Effect on Trait-Unrelated Response Styles](#)

Zachary Loran, UCLA

1:45 PM

[Alternative to ANOVA](#)

Mian Shams Adnan, Bowling Green State University

1:50 PM

[Quantifying the Terms from the Bias-Variance Decomposition Using Metalog Distributions](#)

Neil A Hamlett, Uncertainty Research, LLC

1:55 PM

[Wald Type Tests With the Wrong Dispersion Matrix](#)

Kosman Watte Gedara Dimuthu Hansana Rajapaksha, Southern Illinois University Carbondale

2:00 PM

[The Hierarchy of Block Models](#)

Majid Noroozi, Washington University at St. Louis

2:05 PM

[Fitting Structural Equation Models via Variational Approximations](#)

Khue-Dung Dang, University of Technology Sydney

2:10 PM

[Automatically Extracting Differential Equations from Data with Sparse Regression Techniques](#)

Kevin Egan, Durham University

2:15 PM

[Uniform-Laplace Mixture Distributions](#)

Amos Natido, University of Nevada, Reno

2:20 PM

[Clustering Data with Nonignorable Missingness using Semi-Parametric Mixture Models](#)

Matthieu Marbac, CREST / ENSAI

2:25 PM

[Normal - Pareto Distributions: Theoretical Framework and Computational Issues](#)

Matthew Awere Ohemeng, University of Nevada Reno

2:30 PM

[Scalable Gaussian Processes on Physically Constrained Domains](#)

Bora Jin, Duke University

2:35 PM

[Intrinsic smoothing method based on penalized spherical Bézier curves](#)

Kwan-Young Bak, Korea University

2:40 PM

[Strong Orthogonal Arrays and Orthogonal Array-Based Latin Hypercube Designs for Planning Experiments: A Method for Improving Study Design in Biomedical Research](#)

Kazeem Adewale Osuolale, Nigerian Institute of Medical Research, Yaba, Lagos

2:45 PM

[Multi-scale Affinities with Missing Data: Estimation and Applications](#)

Min Zhang, North Carolina State University

2:50 PM

[Alternatives to Chebyshev's Inequality](#)

Mian Shams Adnan, Bowling Green State University



CS25 - Geo-Spatial Data Analysis

Refereed

Fri, Jun 4, 3:20 PM - 4:55 PM

Chair(s): Brennan Bean, Utah State University

3:25 PM

[Quantifying the Effects of Passenger-Level Heterogeneity on Transit Journey Times](#)

Ramandeep Singh, Imperial College London

3:55 PM

[Evaluating Fairness in the Presence of Spatial Autocorrelation](#)

Subhabrata Majumdar, AT&T Labs - Research

4:25 PM

[Statistical Learning for Predicting Electron Dynamics](#)

Prachi Gupta, University of California, Merced



CS26 - Data-Driven Science

Refereed

Fri, Jun 4, 3:20 PM - 4:55 PM

Chair(s): David Han, The University of Texas at San Antonio

3:25 PM

[Spatial-Temporal Change Detection Using Elastic Functional Data Analysis](#)

Avipsa Roy, Arizona State University

3:55 PM

[Increasing Integration of Data-Driven Analyses in Operational Activities Through Knowledge Management](#)

Thushara Gunda, Sandia National Laboratories

4:25 PM

[Detecting and Explaining Changes in Presidential Approval with Interval-Censored Polling Data](#)

Jiahao Tian, University of Virginia



CS27 - Random Forests Shaping Decisions

Refereed

Fri, Jun 4, 3:20 PM - 4:55 PM

Chair(s): Krithika Suresh, University of Colorado

3:25 PM

[Augmented Bagging and Implications for Variable Importance](#)

Siyu Zhou, University of Pittsburgh

3:55 PM

[Causal Effect Random Forest of Interaction Trees for Learning Individualized Treatment Regimes with Non-Binary Treatments in Observational Studies](#)

Richard A Levine, San Diego State University



CS28 - Honoring the Data Science Accomplishments of William S. Cleveland and John W. Tukey
Special Session

Fri, Jun 4, 3:20 PM - 4:55 PM

Chair(s): Tim Hesterberg, Google

3:25 PM

[Bill Cleveland: Il Maestro of Statistical Graphics](#)

Nicholas Fisher, University of Sydney & ValueMetrics Australia

3:40 PM

[Bill Cleveland's Contributions to Methods for Analysis and Visualization of Large Complex Data](#)

Ryan Hafen, Preva Group

3:55 PM

[Bill Cleveland's Contributions to Analysis of Big Complex Data](#)

Wen-wen Tung, Purdue University

4:10 PM

[Bill Cleveland's Action Plan for Data Science](#)

David Donoho, Stanford University

4:25 PM

[How John Tukey Led Us Into Data Science](#)

William S. Cleveland, Purdue University

GS03 - Closing and Plenary Panel: 2020 Census

General Session

Fri, Jun 4, 5:10 PM - 6:20 PM

Chair(s): Claire Bowen, Urban Institute

[2020 Census](#)

D'Vera Cohn, Pew Research Center; Hansi Lo Wang, NPR; Nancy Potok, NAPx Consulting LLC