Printable Program
Printed on: 12/12/2025

Wednesday, June 3

Welcome & Keynote Address
General Session
Wed, Jun 3, 11:45 AM - 1:00 PM

Chair(s): David Hunter, Penn State University

Instead of Just Teaching Data Science, Let's Understand How and Why People Do It
Rebecca Nugent, Carnegie Mellon University

Computational Statistics Posters
E-Poster
Wed, Jun 3, 1:00 PM - 4:00 PM

Poster Q&A will be available during these designated hours as part of the virtual conference.

1
WITHDRAWN Statistical Computing and Informatics for Biodiversity and Forests Conservation
2
Testing for Heteroscedasticity in Functional Linear Models
James Triece Cameron, George Mason University
3
Comparative Study of Gaussian Stochastic Process Models Under Different Correlation Functions
Kazeem Adewale Osuolale, Nigerian Institute of Medical Research (NIMR)
4
WITHDRAWN The Effect of Institutions on Economic Growth: An Analysis Based on Bayesian Panel Data Estimation
5
WITHDRAWN False Discovery Rates for Second-Generation P-Values in Large-Scale Inference
6
A Moving Shape (3 D) Time Series Model
Mian Arif Shams Adnan, Bowling Green State University
7
Fast Optimal Subsampling Probability Approximation for Generalized Linear Models
JooChul Lee, University of Connecticut, Department of Statistics

Applying Network and Graph Analysis
Invited

Wed, Jun 3, 1:15 PM - 2:50 PM

Organizer(s): Brendan Newlon, Center for Creative Leadership
Chair(s): Sharmistha Guha, Duke University

1:20 PM
Applying Graph Analysis to Explore Thematic Complexity in Qualitative Interview Data
Brendan Newlon, Center for Creative Leadership
1:50 PM
Learning Social Network from Text Data
Xiaoyi Yang, Carnegie Mellon University
2:20 PM
Network Optimization to Evaluate Public Transportation Systems: A Traveling Salesman Problem Example in Paris
Carlos Pinheiro, SAS Institute

Real-Life Data Analysis Experiences for Statistics and Data Science Students

Invited
Wed, Jun 3, 1:15 PM - 2:50 PM

Organizer(s): Nicole Lazar, University of Georgia
Chair(s): David Hunter, Penn State University

1:20 PM
Data Science Clinic: A Capstone Experience for Smithies in SDS
Benjamin S Baumer, Smith College
1:50 PM
Team-Based Learning in a Statistical Consulting Course
John Gabrosek, Grand Valley State University
2:20 PM
A Year-Long Writing-Intensive Capstone in Statistics
Lynne Seymour, University of Georgia

Data Science Using JMP and SAS

Invited
Wed, Jun 3, 1:15 PM - 2:50 PM

Organizer(s): Ruth Hummel, JMP
Chair(s): Robert Winston Blanchard, SAS

1:20 PM
Using JMP for Advanced Data Analytics

Richard D. De Veaux, Williams College
1:50 PM
JMP and SAS – Alone, Together, and with Open-Source Software – for Teaching and Doing Data Science
Ruth Hummel, JMP
2:20 PM
SAS Is Open! SAS as an Open Data Science Platform in the Classroom and on the Job
James Luxton Harroun, SAS Institute

Data Visualization 1

Contributed Refereed
Wed, Jun 3, 1:15 PM - 2:50 PM

Chair(s): Sanvesh Srivastava, University of Iowa

1:20 PM
Data Visualization and Accessibility
Christine P. Chai, Microsoft
1:50 PM
Data Visualization for the Validation of High-Dimensional Data
Aaron Robert Williams, Urban Institute
2:20 PM
Q&A

Machine Learning 1

Contributed Refereed
Wed, Jun 3, 1:15 PM - 2:50 PM

Chair(s): Xiaotong Jiang , University of North Carolina at Chapel Hill

1:20 PM
RankFromSets: Scalable Set Recommendation with Optimal Recall
Jaan Altosaar, Princeton University
1:50 PM
Counterfactual Demand Predictions with Deep Learning
Mingyu (Max) Joo, UC Riverside
2:20 PM
Explaining the Practical Success of Random Forests
Siyu Zhou, University of Pittsburgh

Best Practices for Leading DS Efforts in Your Organization

Panel Discussion
Wed, Jun 3, 3:00 PM - 4:30 PM

Chair(s): David Hunter, Penn State University

We will focus on how data science groups are organized within data science institutes in academia and within industry and government.

3:05 PM
Best Practices for Leading DS Efforts in Your Organization
Jeannette M. Wing, Columbia University; Rebecca Nugent, Carnegie Mellon University; Claire McKay Bowen, Urban Institute; Erin LeDell, H2O.ai

SC1 SOLD OUT - Big Data, Data Science, and Deep Learning for Statisticians, Part 1 (Ticket Required)
Short Course
Wed, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Ming Li, Amazon

With the recent big data, data science, and deep learning revolution, companies across the world are hungry for data scientists and machine learning scientists to bring actionable insight from the vast amount of data collected. In the past couple of years, deep learning has gained traction in many application areas and become an essential tool in the data scientist’s toolbox.

In this course, participants will develop a clear understanding of the big data cloud platform and technical skills in data science and machine learning. They will use hands-on exercises to understand deep learning. We will also cover the “art” part of data science and machine learning so participants learn the typical data science project flow, general pitfalls in data science and machine learning, and soft skills to effectively communicate with business stakeholders.

The big data platform, data science, and deep learning overviews are specifically designed for an audience with a statistics education background. This course will prepare statisticians to be successful data scientists and machine learning scientists in various industries and business sectors with deep learning as a focus. Please have a laptop available for hands-on sessions. No software download or installation is needed.

SC2 - Introduction to Programming Quantum Computers (Ticket Required)
Short Course
Wed, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Mark Fingerhuth, Quantum Open Source Foundation and ProteinQure

Quantum computing isn’t science fiction anymore. IBM, D-Wave, and Rigetti all provide cloud access to their quantum processing units (QPUs). Have a laptop available! We will talk about the basics of quantum computing and how to implement an algorithm on actual quantum hardware. We will focus on Rigetti’s Forest SDK, a set of Python libraries designed to interact with QPU, and practical quantum computing, rather than theory. Participants will learn about the following:

• The notion of a quantum bit • Different quantum computing architectures • Various quantum logic operations and how to implement them in code • Rigetti's Python API to interact with the quantum device • How to write and execute a quantum program

SC3 - How to Create a Development Environment for Reproducible Research (Ticket Required)

Short Course
Wed, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Brian Lee Yung Rowe, Pez.AI

Winston Churchill observed that "we shape our buildings, and afterwards our buildings shape us." The same is true of our development environment, which shapes our development process. Ad hoc and unstructured environments lead to unstructured processes that are difficult to reproduce. This short course leverages the author's crant toolkit and shows how to use Docker, git, make, and other tools to create a development environment optimized for reproducible research. At the end of the course, you'll be able to create a re-usable environment that automates testing, packaging, report generation, and more. You'll also learn how to incorporate notebooks into your development process in a way that maintains reproducible research.

SC4 - Recommendation Systems and Reinforcement Learning for Data Scientists (Ticket Required)

Short Course
Wed, Jun 3, 3:00 PM - 6:30 PM

Instructor(s): Ying Lu, Google; Wutao Wei, Twitter

We all hear about data science technology these years. What is data science? How does data science change the things around us? This short course serves as an introduction to a combination of practical data science technologies with a focus on experimentation, recommendation systems, and reinforcement learning. We will talk about how these core technologies help build a great product. At the end of the course, audience is expected have a clear understanding of various data science technologies and applications. Both lectures and lab exercises will be offered.

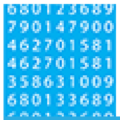
Thursday, June 4

Anomaly Detection in Complex Data

Invited
Thu, Jun 4, 10:00 AM - 11:35 AM

Organizer(s): Sarah Rajtmajer, Penn State University
Chair(s): Sarah Rajtmajer, Penn State University

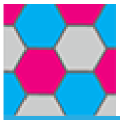
10:05 AM
High Temperature Structure Detection in Ferromagnets
Matey Neykov, University of Pittsburgh
10:35 AM
Toward Secure and Interpretable AI: Scalable Methods, Interactive Visualizations, and Practical Tools
Polo Chau, Georgia Tech
11:05 AM
Detecting Anomalies in Graph-Structured Data
James Sharpnack, UC Davis



Data Science Using R
Invited
Thu, Jun 4, 10:00 AM - 11:35 AM

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

10:05 AM
Bayesian Methods for Data Science Using R
Christina Knudson, University of St. Thomas
10:35 AM
Process Automation as the Backbone of Reproducible Science
Brian Lee Yung Rowe, Pez.AI
11:05 AM
Training Large Deep Learning Models Using Spark, TensorFlow, and R
Javier Luraschi, RStudio



Computing in Data Privacy
Invited
Thu, Jun 4, 10:00 AM - 11:35 AM

Organizer(s): Aleksandra Slavkovic, Penn State University
Chair(s): Aleksandra Slavkovic, Penn State University

10:05 AM
Formally Private Microdata at Scale: Reducing the Magnitude of Upward Bias
Philip Leclerc, United States Census Bureau

10:35 AM
OpenDP: An Open-Source Suite of Differential Privacy Tools
James Honaker, Harvard University
11:05 AM
Encode, Shuffle, Analyze Revisited: Strong Privacy Despite High Epsilon
Abhradeep Guha Thakurta, Google Research Brain Team and UC Santa Cruz

Visualization for Big Data and AI
Invited
Thu, Jun 4, 10:00 AM - 11:35 AM

Organizer(s): Andee Kaplan, Colorado State
Chair(s): Haley Jeppson, Iowa State University

10:05 AM
Telling a Visual Story Within Big Data: Case Studies on Interactive Visualizations for Supercomputer Data
Claire McKay Bowen, Urban Institute
10:35 AM
Protoshiny: Interactive Exploration of Dendrograms with Prototypes
Jacob Bien, University of Southern California
11:05 AM
Visualizing Complex Science
Samuel F. Way, Spotify

Education 1
Contributed Refereed
Thu, Jun 4, 10:00 AM - 11:35 AM

Chair(s): Donna LaLonde, American Statistical Association

10:05 AM
Teaching the Gestalt Principles to Help Undergraduate Students Design Effective Tables and Graphs
Silas Bergen, Winona State University
10:20 AM
Bringing Visual Inference to the Classroom
Adam Loy, Carleton College
10:35 AM
Data Management with Data Verbs
Todd Iverson, Winona State University
10:50 AM
Beyond NYC Flights in Intro to Data Science: Curtis Flowers and the Role of Race in Jury Selection
Paul Roback, St. Olaf College

11:05 AM
Q&A

Practice and Applications 1

Contributed Refereed
Thu, Jun 4, 10:00 AM - 11:35 AM

Chair(s): David Hunter, Penn State University

10:05 AM
Leveraging Methods for Subsampling: Toward a Realistic Evaluation
Changrui Liu, University of Kentucky
10:35 AM
Improving Cloud Infrastructure Capacity Planning Decisions with Scalable Human-in-the-loop Scenario Forecasting
Jiaping Zhang, Salesforce
11:05 AM
Finite Sample Properties of an Exponential-Compound Symmetric Covariance Structure
Amber K. Weydert, University of West Florida

Education and Data Visualization Posters

E-Poster
Thu, Jun 4, 10:00 AM - 1:00 PM

Poster Q&A will be available during these designated hours as part of the virtual conference.

1
Exploring Technical Competencies Needs for Future Information Technology Workforce
Ana Valentin, Marymount University
2
WITHDRAWN The Arcus Learning Exchange: Cross-Departmental Education Development at the Children's Hospital of Philadelphia
3
Increasing Diversity in Biomedical Data Science: Implementation and Impact of Best Practices
Judith E Canner, California State University, Monterey Bay
4
ACM Draft 2 Computing Competencies for Undergraduate Data Science
Karl Schmitt, Valparaiso University
5
A Statistician Teaches Deep Learning: From Fundamentals to Applications

David Han, The University of Texas at San Antonio

6

Identifying Academic At-Risk Students with Consistence Validation Using Predictive Analytics

Jianbin Zhu, University of Central Florida

7

Educational Tool and Active-Learning Class Activity for Teaching Agglomerative Hierarchical Clustering

Xizhen Cai, Williams College

8

REDCap and RShiny Together to Survey and Deliver Personalized Feedback of a Well-Being Assessment

Duncan Grade Vos, WMU School of Medicine

9

Modified Box Plots for Arithmetic, Geometric, and Harmonic Observations

Mian Arif Shams Adnan, Bowling Green State University

10

Geometries of the Connections of the Graphical Presentations of Several Statistical Tools

Mian Arif Shams Adnan, Bowling Green State University

11

CatViz for Visual Exploration of High-Dimensional Categorical Data Sets

Raif Rustamov, AT&T Labs Research

12

A Range-Based Box Plot

Mian Arif Shams Adnan, Bowling Green State University

13

Building an Open-Sourced Geospatial Visualization Shiny Application in R for Healthcare Providers and Evaluators

Dar'ya Y Pozhidayeva, Oregon Health & Science University

Parallel Computing

Invited

Thu, Jun 4, 11:40 AM - 12:45 PM

Organizer(s): Sean Blanchard, Los Alamos National Laboratory
Chair(s): Sean Blanchard, Los Alamos National Laboratory

11:45 AM

Democratizing Calculations in the Cloud

Andrew Glenn Shewmaker, OpenEye Scientific

12:15 PM

Adaptive MCMC for Everyone

Jeffrey S. Rosenthal, University of Toronto

Harnessing the Power of Data to Promote Institutional Change at Higher Education Institutions

Invited
Thu, Jun 4, 11:40 AM - 12:45 PM

Organizer(s): Kameryn Denaro, UC Irvine
Chair(s): Wendy Martinez, Bureau of Labor Statistics

11:45 AM
Making an Impact in an Institutional Research Office: On Data Champions and Machine Learning
Richard A. Levine, San Diego State University
12:15 PM
A Data-Driven Approach to Promoting Innovation and Excellence in Teaching at Higher Education Institutions
Kameryn Denaro, UC Irvine

Modern Inference in Statistical Machine Learning

Invited
Thu, Jun 4, 11:40 AM - 12:45 PM

Organizer(s): Ryan Tibshirani, Carnegie Mellon University
Chair(s): Nicholas Schmidt, BLDS

11:45 AM
Predictive Inference with Random Forests
Lucas Mentch, University of Pittsburgh
12:15 PM
Semiparametric Estimation in High Dimensions
Mladen Kolar, U Chicago Booth

Machine Learning 5

Contributed Refereed
Thu, Jun 4, 11:40 AM - 12:45 PM

Chair(s): Thomas Carpenito, Northeastern University

11:45 AM
Modernizing k-Nearest Neighbors Software
Norm Matloff, UC Davis
12:15 PM
Heterogeneous Treatment Effects of Medicaid and Efficient Policies
Shishir Shakya, West Virginia University

Machine Learning 6

Contributed Refereed

Thu, Jun 4, 11:40 AM - 12:45 PM

Chair(s): Yirui Hu, Geisinger

11:45 AM

Functional Singular Spectrum Analysis

Mehdi Maadooliat, Department of MSSC at Marquette University

12:15 PM

Statistical Learning and Energy Statistics for High-Dimensional Time Series

John Steven Schuler, George Mason University

Practice and Applications 5

Contributed Refereed

Thu, Jun 4, 11:40 AM - 12:45 PM

Chair(s): Lauren Alpert Sugden, Duquesne University

11:45 AM

Estimation Graphics: Essential Data Analysis for Biomedical Science

Joses Ho, Institute for Molecular and Cell Biology

12:00 PM

Trial-by-Trial Mid-Frontal Theta Power Predicts Emotional Decision Processes in Response Inhibition Task

Siddharth Nayak, Institute of Statistical Science, Academia Sinica

12:15 PM

A Paradigm for Managing Computational Reproducibility in a Changing Software Package Landscape

Kiegan Rice, Iowa State University

Divide and Recombine for Big Data Analysis and Visualization

Invited

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

Organizer(s): Susan Vanderplas, Iowa State University
Chair(s): Susan Vanderplas, Iowa State University

1:25 PM

Divide and Recombine (D&R) with R-RHIPE-Hadoop Software
William S. Cleveland, Purdue University
1:55 PM
Rethinking Climate Data Analysis and Visualization in the Era of Big Data
Wen-wen Tung, Purdue University
2:25 PM
Distributed Bayesian Varying Coefficient Modeling Using a Gaussian Process Prior
Sanvesh Srivastava, University of Iowa

Interactive Machine Learning
Invited
Thu, Jun 4, 1:20 PM - 2:55 PM

Organizer(s): James Sharpnack, UC Davis
Chair(s): James Sharpnack, UC Davis

1:25 PM
On the Global Convergence of Policy Optimization in Deep Reinforcement Learning
Zhaoran Wang, Northwestern University
1:55 PM
WITHDRAWN: Marginal Posterior Sampling for Slate Bandits
2:25 PM
Interactive Learning Using Labels and Comparisons
Aarti Singh, Carnegie Mellon University

Community Engagement Through Data Science Education
Invited
Thu, Jun 4, 1:20 PM - 2:55 PM

Organizer(s): Leah Jager, Johns Hopkins Bloomberg School of Public Health
Chair(s): Leah Jager, Johns Hopkins Bloomberg School of Public Health

1:25 PM
Can Data Science Education Be Used as a Tool for Upward Mobility?
Aboozar Hadavand, Johns Hopkins University, Bloomberg School of Public Health
1:55 PM
Incorporating Community-Based Learning Into the Classroom
Lynne Steuerle Schofield, Swarthmore College
2:25 PM
Statistics in the Community: Community-University Partnerships Fostering Data Science Education
Stephen Salerno, Department of Biostatistics, University of Michigan

Cloud Computing: The Future for Data Science Applications
Invited
Thu, Jun 4, 1:20 PM - 2:55 PM

Organizer(s): Ming Li, Amazon
Chair(s): Ruth Hummel, JMP

1:25 PM
End-to-End Data Science Project Cycle
Ming Li, Amazon
1:55 PM
Machine Learning and Cloud Computing for Statisticians
Robert Winston Blanchard, SAS
2:25 PM
Q&A

Computational Statistics 1
Contributed Refereed
Thu, Jun 4, 1:20 PM - 2:55 PM

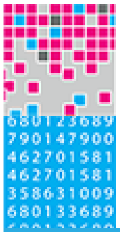
Chair(s): Sujay Datta, University of Akron

1:25 PM
Nonparametric Estimation of Blood Alcohol Concentration from Transdermal Alcohol Measurements Using Alcohol Biosensor Devices
Bryan Edward Vader, Naval Base Ventura County
1:55 PM
Parameter-Expanded Data Augmentation for Analyzing Correlated Binary Data Using Multivariate Probit Models
Xiao Zhang, Michigan Technological University
2:25 PM
Streaming Data Analysis with Dynamic Regression Trees
Simon Paul Wilson, Trinity College Dublin
2:40 PM
Q&A

Practice and Applications 2
Contributed Refereed
Thu, Jun 4, 1:20 PM - 2:55 PM

Chair(s): Mitra Devkota, University of North Georgia

1:25 PM
Scaleable Correlated Topic Modelling for Job Matching
Simon Paul Wilson, Trinity College Dublin
1:40 PM
Bayesian Inference for Polycrystalline Materials
James Matuk, The Ohio State University
1:55 PM
SVM Model for Blood Cell Classification Using Interpretable Features Outperforms CNN-Based Approaches
William Franz Lamberti, George Mason University
2:10 PM
Visualizing the Food Landscape of Durham with Tableau
Joseph Lewis Graves, NCAT
2:25 PM
A Spatiotemporal Case Crossover Model of Asthma Attacks in the City of Houston
Julia Schedler, Rice University
2:40 PM
Q&A



Machine Learning and Software and Data Science Technologies Posters
E-Poster
Thu, Jun 4, 2:00 PM - 5:00 PM

Poster Q&A will be available during these designated hours as part of the virtual conference.

1
WITHDRAWN: Prediction of Hospital Readmissions: A Comparison of Predictive Methods on Binary and Survival Outcomes
2
WITHDRAWN Prediction of Inpatient Quality Indicators: A Comparison of Predictive Methods with and Without Random Hospital Effect
3
Can Big Data Algorithms Be Used to Improve Cybersecurity?
Allen Sina Rahrooh, University of Central Florida
4
WITHDRAWN: BLNN: An R Package for Training Neural Networks
5
WITHDRAWN Decision Tree Model-Based Gene Selection and Classification for Breast Cancer Risk Prediction
6
Learning the Stock Market States via a Logistic Regression Model and Its Applications

Qiyu Wang, Zhejiang Univ of Finance and Econ
7
Multiple Sequence Alignment Using Tensor Analysis
Mian Arif Shams Adnan, Bowling Green State University
8
Investigation of the Interplay Between Random Forest and Kernel Methods in Big Data
Richard Baumgartner, Merck&Co., Inc.
9
Interfacing Statistical Software Packages with R and Python
Neil Polhemus, Statgraphics Technologies, Inc.
10
TF-IDF-Weighted Similarity Estimates for Unseen Categories
Handong David Bang, UNC Chapel Hill Department of Biostatistics
11
Developing a Computational Framework for Precise TAD Boundary Prediction Using Genomic Elements
Spiro C Stilianoudakis, Virginia Commonwealth University
12
Predicting 30-Day Readmission After Surgery Among Colorectal Cancer Patients
Anshul Saxena, Baptist Health south Florida
13
R Package mase
Iris Griffith, Reed College

Ethics and Bias in Algorithms
Panel Discussion
Thu, Jun 4, 3:00 PM - 4:30 PM

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

In this mini-workshop, we discuss some of the social and ethical challenges of statistical and machine learning algorithms with a panel of experts from academia and industry.

3:05 PM
Ethics and Bias in Algorithms
Jie Chen, Wells Fargo; Jim Rosenberger, NISS; Aleksandra Slavkovic, Penn State University; Robert Tibshirani, Stanford University

SC1 SOLD OUT - Big Data, Data Science, and Deep Learning for Statisticians, Part 2 (Ticket Required)
Short Course
Thu, Jun 4, 3:00 PM - 6:30 PM

Instructor(s): Ming Li, Amazon

Continuation of course.

SC5 - CANCELLED: Building Advanced Computer Vision Models Using SAS Software (Ticket Required)
Short Course
Thu, Jun 4, 3:00 PM - 6:30 PM

SC6 - Data Science Workflows Using R and Spark (Ticket Required)
Short Course
Thu, Jun 4, 3:00 PM - 6:30 PM

Instructor(s): Jim Harner, West Virginia University

R is a flexible, extensible statistical computing environment, but it is limited to single-core execution. Spark is a distributed computing environment that treats R as a first-class programming language. This course introduces data structures in R and their use in functional programming workflows relevant to data science.

The course covers the initial steps in the data science process: - extracting data from source systems, - transforming data into a tidy form, - loading data into distributed file systems, distributed data warehouses, and NoSQL databases, i.e., ETL.

These R-based workflows are illustrated by using dplyr directly and as a frontend to SQL databases. The sparklyr package with its dplyr interface to Spark is then used for modeling big data using regression and classification supervised learning methods. Unsupervised learning methods, such as clustering and dimension reduction, are also covered. Finally, methods for analyzing streaming data are presented. Student accounts are provided to allow attendees to interactively run the R Markdown content in Amazon’s cloud (AWS). The computing infrastructure and the content is containerized which allows the complete course environment to be downloaded and run on Docker-supported laptops.

SC7 - Visualizing Big Data (Ticket Required)
Short Course
Thu, Jun 4, 3:00 PM - 5:00 PM

Instructor(s): Leland Wilkinson, H2O.ai and University of Illinois at Chicago

Big datasets (many rows, many columns, many items, ...) present special problems for visualization. Even when trying to plot simple rectangular datasets, we encounter complexity (many functions are polynomial or exponential in rows or columns), the curse of dimensionality (distances approach a constant as dimensionality heads toward infinity), choke points (data bus or network bandwidth), and limited display resolution (even with megapixel displays). This workshop covers recent strategies that exploit aggregation and projection to reduce datasets to manageable proportions. It also covers graphic representations that are most suitable for exploring multivariate data.

Friday, June 5

Keynote Address
General Session
Fri, Jun 5, 10:00 AM - 11:10 AM

Chair(s): David Hunter, Penn State University

Data for Good: Ensuring the Responsible Use of Data to Benefit Society
Jeannette M. Wing, Columbia University

Practice and Applications Posters, Part 1
E-Poster
Fri, Jun 5, 10:00 AM - 1:00 PM

Poster Q&A will be available during these designated hours as part of the virtual conference.

1
How Does Mental Health Affect Unemployment? The Mediation Effect of Concentration Ability
Chuhan Ouyang, 2003
2
Two Notes About the Two Faces of R-Squared
Gyasi K Dapaa, Indeed Inc
3
The Children's National Data Lake (CNDL): A Partnership with Amazon Web Services and Cerner
James Bost, Children's National Hospital
4
PM2.5 Data from Functional Time Series Analysis
5
Tornado: Classification, Correlation, Prediction
Thilini Vasana Mahanama, Texas Tech University
6
Detecting Cell Culture Contamination with Multivariate Time Series Data
Laura L Tupper, Williams College
7
A Novel Application of Time Series Classification Using the Continuous Wavelet Transform and Convolutional Neural Networks on Smartphone Sensor Data
William Robert Nadolski, SAS Institute
8
Application of Unsupervised Machine Learning Technique in Early Discovery Efforts to Identify Novel Subgroups of Patients with Type 2 Diabetes Mellitus Using Proinsulin Levels

Santosh C Sutradhar, Merck & Co., Inc.
9
Statistical Modeling of Emission Factors of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in Africa
Mohamed Ali Abu Sheha, University of South Florida
10
Independence Test for Bivariate Pareto Data
William Cipolli, Colgate University
11
Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects: An Empirical Application
Shaymal Chandra Halder, Auburn University
12
A Lattice and Random Intermediate Point Sampling Design for Animal Movement
Elizabeth Eisenhauer, Pennsylvania State University
13
Data Science at the Mayo Clinic: Implementation of the Discovery, Translation, and Application (DTA) Framework in Outpatient Palliative Care Practice
Shusaku William Asai, Mayo Clinic

Recent Advances in Entity Resolution
Invited
Fri, Jun 5, 11:15 AM - 12:50 PM

Organizer(s): Rebecca Steorts, Duke University
Chair(s): Christine P. Chai, Microsoft

This session will not be recorded/available for registrants to access on-demand.

11:20 AM
Challenges for Accurate Enumerations of Census and Voter Registrations Databases
Rebecca Steorts, Duke University
11:50 AM
Bayesian Canonicalization of Voter Registration Files
Andee Kaplan, Colorado State
12:20 PM
Multifile Record Linkage and Duplicate Detection via a Structured Prior for Partitions
Serge Aleshin-Guendel, University of Washington

Interpretable and Fair Machine Learning in Finance
Invited
Fri, Jun 5, 11:15 AM - 12:50 PM

Organizer(s): Patrick Hall, H2O.ai
Chair(s): Patrick Hall, H2O.ai

This session will not be recorded/available for registrants to access on-demand.

11:20 AM
Adaptive Explainable Neural Networks (AxNN)
Jie Chen, Wells Fargo
11:50 AM
Responsible Data Science: Identifying and Fixing Biased AI
Nicholas Schmidt, BLDS
12:20 PM
A New Approach to Providing Explanations for Machine Learning Algorithms
Tom Prendergast, Synchrony Financial

Statistics for the Engaged Citizen: Revising Educational Practices to Increase Relevance in Everyday Life

Invited
Fri, Jun 5, 11:15 AM - 12:50 PM

Organizer(s): Leslie Myint, Macalester College
Chair(s): Leslie Myint, Macalester College

11:20 AM
Best Practices for Teaching R Programming to Students: A Randomized Controlled Trial
Lucy D'Agostino McGowan, Wake Forest University
11:50 AM
Real Data Analysis in the Classroom Through the Use of Case Studies
Leah Jager, Johns Hopkins Bloomberg School of Public Health
12:20 PM
Q&A

Data Journalism and Visualization

Invited
Fri, Jun 5, 11:15 AM - 12:50 PM

Organizer(s): Kiegan Rice, Iowa State University
Chair(s): Kiegan Rice, Iowa State University

11:20 AM
How (and Why) Election Results Data Gets Made
Derek Willis, OpenElections
11:50 AM
Designing Information Graphics for Communication
Peter Bell, Pew Research Center

12:20 PM
Design for Journalism
Sarah Almukhtar, The New York Times

Computational Statistics 2

Contributed Refereed
Fri, Jun 5, 11:15 AM - 12:50 PM

Chair(s): Waldyn Gerardo Martinez, Miami University

11:20 AM
On the Estimation Bias in First-Order Bifurcating Autoregressive Models
Tamer Elbayoumi, North Carolina A&T State University
11:50 AM
Learning Large Genetic Networks Using Gaussian Graphical Models
Sujay Datta, University of Akron
12:20 PM
Kernel Mean Embedding-Based Hypothesis Tests for Comparing Spatial Point Patterns
Raif Rustamov, AT&T Labs Research

Software and Data Science Technologies 1

Contributed Refereed
Fri, Jun 5, 11:15 AM - 12:50 PM

Chair(s): Jim Harner, West Virginia University

11:20 AM
Creating Optimal Conditions for Reproducible Data Analysis in R with `Fertile`
Benjamin S Baumer, Smith College
11:35 AM
Likelihood-Based Inference for Generalized Linear Mixed Models: Inference with R Package glmm
Christina Knudson, University of St. Thomas
11:50 AM
Q&A

Bits and Bytes Networking Break

Social Event
Fri, Jun 5, 12:50 PM - 1:20 PM

Grab a bite to eat and connect with fellow attendees for conversation.

Interactive Graphics

Invited
Fri, Jun 5, 1:25 PM - 3:00 PM

Organizer(s): Andee Kaplan, Colorado State
Chair(s): Alex Kale, University of Washington

This session will not be recorded/available for registrants to access on-demand.

1:30 PM
Reproducible Shiny Apps with Shinymeta
Carson Sievert, RStudio
2:00 PM
Vega-Lite: What Does a Grammar of Interactive Graphics Enable?
Arvind Satyanarayan, MIT CSAIL
2:30 PM
Connecting HTML Widgets with Shiny-Like Inputs to Visualize the Structure of High-Dimensional Data Using Tours
Haley Jeppson, Iowa State University

Bayesian Computation

Invited
Fri, Jun 5, 1:25 PM - 3:00 PM

Organizer(s): Michele Guindani, University of California, Irvine
Chair(s): Michele Guindani, University of California, Irvine

1:30 PM
Convergence Analysis of a Collapsed Gibbs Sampler for Bayesian Vector Autoregressions
Galin Jones, University of Minnesota
2:00 PM
Robust, Efficient Hamiltonian Monte Carlo Algorithms on Manifolds
Shiwei Lan, Arizona State University
2:30 PM
Q&A

Data Science in Industry

Invited
Fri, Jun 5, 1:25 PM - 3:00 PM

Organizer(s): Philip Turk, Western Data Analytics, LLC

Chair(s): Soren Harner, LayerJot

This session will not be recorded/available for registrants to access on-demand.

1:30 PM
Safely Self-Driving at Scale
Nicholas Armstrong-Crews, Waymo
2:00 PM
Leveraging Data Science to Support Clinical Trial Execution
Matthew Austin, Amgen, Inc
2:30 PM
Row Crop Breeding at a Global Scale: Applications of Software and Decision-Science at Bayer Crop Science
Ross S. Bricklemyer, Bayer Crop Science

Machine Learning 2
Contributed Refereed
Fri, Jun 5, 1:25 PM - 3:00 PM

Chair(s): Lynne Steuerle Schofield, Swarthmore College

This session will not be recorded/available for registrants to access on-demand.

1:30 PM
Deep Doubly Robust Outcome-Weighted Learning
Xiaotong Jiang, University of North Carolina at Chapel Hill
2:00 PM
Locally Optimized Random Forests: A Solution to Forecasting Severe Hurricane Power Outages
Tim Coleman, University of Pittsburgh, Department of Statistics
2:30 PM
Modern Multiple Imputation Applied to Functional Data
Aniruddha Rajendra Rao, Pennsylvania State University

Machine Learning 4
Contributed Refereed
Fri, Jun 5, 1:25 PM - 3:00 PM

Chair(s): Andee Kaplan, Colorado State

1:30 PM
A General Framework for Empirical Bayes Estimation in Discrete Linear Exponential Family
Trambak Banerjee, University of Southern California
2:00 PM
Multivariate Functional Singular Spectrum Analysis Over Different Dimensional Domains

Jordan Christopher Trinka, Department of MSSC at Marquette Univeristy
2:30 PM
A One-Class Peeling Method for Anomaly Detection
Waldyn Gerardo Martinez, Miami University

Practice and Applications 3
Contributed Refereed
Fri, Jun 5, 1:25 PM - 3:00 PM

Chair(s): William Franz Lamberti, George Mason University

1:30 PM
Statistical Inference of Adaptive Mutations and Genes from Worldwide Genome Sequences
Lauren Alpert Sugden, Duquesne University
2:00 PM
Q&A

Practice and Applications Posters, Part 2
E-Poster
Fri, Jun 5, 2:00 PM - 5:00 PM

Poster Q&A will be available during these designated hours as part of the virtual conference.

1
On Using Graphical Models and Regularized Parameter Estimates: Practical Considerations and Applications
Zhipu Zhou, University of California - Santa Barbara
2
Development of an Integrated Oncology Data Warehouse for Data Science and Precision Medicine Applications to Facilitate Complex Clinical Decisions
Anshul Saxena, Baptist Health south Florida
3
Data-Driven Statistical Modeling and Analysis of the Survival Times of Multiple Myeloma Cancer
Lohuwa Mamudu, University of South Florida
4
Predicting International Conflict Onset
Daniel Kent, The Ohio State University
5
Optimal Dynamic Treatment Regime by Reinforcement Learning in Clinical Medicine
David Han, The University of Texas at San Antonio
6
Prediction and Modeling of Sensor Endpoint Data in Clinical Trials
Yi-Ting Chang, AstraZeneca

7
Associations Between Accelerometry-Based Gait Measures and Life-Space Assessment Scores in Older Adults
Anisha Suri, University of Pittsburgh

8
A Selective Inference Approach for FDR Control Using Multi-Omics Covariates Yields Insights into Disease Risk
Ronald Yurko, Carnegie Mellon University

9
WITHDRAWN: Basket Analysis Methods in Big Data: The Case of Diabetes

10
A Neural Network Approach for Imputing Missing Metabolomics Data
Tyler Cook, University of Central Oklahoma

11
Statistical Downscaling of Climate-Model Produced Daily Precipitation Based on a Single-Stage Zero-Inflation Rainfall Model
Yiming Liu, University of New Hampshire

12
High-Frequency Multivariate Environmental Time Series: GAM Gap Filling and Distributed Nonlinear Lag Modeling
Lin Wang, University of New Hampshire

13
Solving Data Science Problems in the US Federal Government with R Shiny
Samantha Tyner, U.S. Bureau of Labor Statistics

14
A Survey of Statistical Methods for Investigating Risk of Low-Back Pain in a Cohort of Manufacturing Workers
Charles Ingulli, American University

15
WITHDRAWN Profile of Hospital Admissions due to Asthma: 2012--2017

16
Using Heterogeneous Treatment Effects to Evaluate the Impact of Heath Management Interventions A Simulation Study Using Medical Claims Data
Khalil Zlaoui, Aetna

17
An R Markdown Template for CMS Statistical Reports: The Labyrinth of R Markdown and Microsoft Word
Carina Spicer, Merck

BYOC (Bring Your Own Coffee) Break
Social Event
Fri, Jun 5, 3:00 PM - 3:30 PM

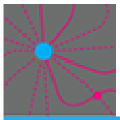
Grab a fix and join your colleagues for a breather between sessions.

R vs. Python, or Both?

Invited
Fri, Jun 5, 3:30 PM - 5:05 PM

Organizer(s): Philip Turk, Western Data Analytics, LLC
Chair(s): Donna LaLonde, American Statistical Association

3:35 PM
R vs. Python for Data Science?
Norm Matloff, UC Davis
4:05 PM
Python and R, When in Rome
Soren Harner, LayerJot
4:35 PM
Q&A

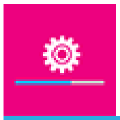


User Testing Statistical Graphics

Invited
Fri, Jun 5, 3:30 PM - 5:05 PM

Organizer(s): Adam Loy, Carleton College
Chair(s): Adam Loy, Carleton College

3:35 PM
Expect Users to Satisfice: Designing Interfaces for Reasoning with Uncertainty
Alex Kale, University of Washington
4:05 PM
Open Data Visualizations and Analytics as Tools for Policy-Making
Loni Hagen, University of South Florida
4:35 PM
Scenarios of Visual Inference
Heike Hofmann, Iowa State University



The Hidden AI Threats: Data, Stability, and Model Decay

Invited
Fri, Jun 5, 3:30 PM - 5:05 PM

Organizer(s): Celeste Fralick, Mcafee
Chair(s): Sarah Kalicin, Intel Corporation

3:35 PM
The Hidden Threats of Decay in AI
Celeste Fralick, Mcafee

4:05 PM
If Your Data Is Bad, Your AI Initiatives Are Doomed!
Thomas C. Redman, "the Data Doc"
4:35 PM
Veridical Data Science and the PCS Framework
Raaz Dwivedi, UC Berkeley

Practice and Applications 4

Contributed Refereed
Fri, Jun 5, 3:30 PM - 5:05 PM

Chair(s): Christina Knudson, University of St. Thomas

3:35 PM
Application of Inverse Probability Weights in a Generalized Linear Mixed Model with Random Intercept to Estimate Causal Treatment Effects in Observational Studies
Duncan Grade Vos, WMU School of Medicine
3:50 PM
Predicting the Lifespan of Drosophila Melanogaster: A Novel Application of Convolutional Neural Networks and Zero-Inflated Autoregressive Conditional Poisson Models
Yi Zhang, Missouri University of Science and Technology
4:05 PM
Bayesian Methods for Estimating the Population Attributable Risk in the Presence of Risk Factor Misclassification
Benedict Wong, Food and Drug Administration
4:20 PM
Bayesian Regression with Undirected Network Predictors with an Application to Brain Connectome Data
Sharmistha Guha, Duke University
4:35 PM
Q&A

Education 2

Contributed Refereed
Fri, Jun 5, 3:30 PM - 5:05 PM

Chair(s): Zhi Yang, USC

3:35 PM
Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years
Aimee Schwab-McCoy, Creighton University
4:05 PM
Hosting a Data Science Hackathon with Limited Resources

Christopher Wedrychowicz, Saint Mary's College

4:35 PM

Q&A

Machine Learning 3

Contributed Refereed

Fri, Jun 5, 3:30 PM - 5:05 PM

Chair(s): Claire McKay Bowen, Urban Institute

3:35 PM

Machine Learning Model Selection with Complex Sample Survey Data

Brian Kim, University of Maryland

3:50 PM

Permutation-Based Uncertainty Quantification

Vaidehi Ulhas Dixit, North Carolina State University

4:05 PM

Toward Sequential Data Clustering via Long Short-Term Memory Auto-Encoder

Yirui Hu, Geisinger

4:20 PM

MISL: Multiple Imputation by Super Learning

Thomas Carpenito, Northeastern University

4:35 PM

Self-Supervised Learning for Outlier Detection

Jan Diers, Friedrich-Schiller-University Jena

4:50 PM

Q&A

Closing Keynote Address

General Session

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

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

Data Science, Statistics, and Health

Robert Tibshirani, Stanford University

Virtual Happy Hour

Social Event

Fri, Jun 5, 6:30 PM - 7:30 PM

Celebrate a job well done with your friends and colleagues by grabbing a congratulatory drink and joining them for an hour of happiness. Rooms will be shuffled, so never a dull moment!