Printable Program
Printed on: 02/14/2026

Wednesday, May 29

Registration
SDSS Hours
Wed, May 29, 7:00 AM - 6:30 PM
Grand Ballroom Foyer

SC1 - Welcome to the Tidyverse: An Introduction to R for Data Science
Short Course
Wed, May 29, 8:00 AM - 5:30 PM
Grand Ballroom E

Instructor(s): Garrett Grolemund, RStudio

Looking for an effective way to learn R? This one day course will teach you a workflow for doing data science with the R language. It focuses on using R's Tidyverse, which is a core set of R packages that are known for their impressive performance and ease of use. We will focus on doing data science, not programming. You'll learn to:

* Visualize data with R's ggplot2 package * Wrangle data with R's dplyr package * Fit models with base R, and * Document your work reproducibly with R Markdown

Along the way, you will practice using R's syntax, gaining comfort with R through many exercises and examples. Bring your laptop! The workshop will be taught by Garrett Grolemund, an award winning instructor and the co-author of _R for Data Science_.

SC2 - Modeling in the Tidyverse
Short Course
Wed, May 29, 8:00 AM - 5:30 PM
Grand Ballroom F

Instructor(s): Max Kuhn, RStudio

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures. In the last two years, a suite of tidyverse packages have been created that focus on modeling. This course walks through the process of modeling data using these tools. A focus is on modeling for prediction and inference as well as feature engineering.

SC3 - Data Visualization: Principles and Applications in R, Tableau, and Python
Short Course
Wed, May 29, 8:00 AM - 12:00 PM
Grand Ballroom G

Instructor(s): Silas Bergen, Winona State University; Todd Iverson, Winona State University

In this course, participants will be introduced to principles of data visualization from foundational literature and implement these principles with hands-on activities using Tableau Public, Python (Altair), and R (ggplot). The course instructors have experience teaching these concepts and content as part of undergraduate statistics and data science curricula, and will use example class projects from these courses. The course will be divided into two modules. Module 1 will cover the principles of data visualization theory, summarizing and illustrating foundational data visualization literature. Module 2 will demonstrate how these principles are applied in various software platforms. Hands-on data visualization tasks will be employed throughout. Participants must bring their own laptops.

SC4 - Reproducible Research with R
Short Course
Wed, May 29, 8:00 AM - 12:00 PM
Grand Ballroom I

Instructor(s): Kara Woo, Sage Bionetworks

This course will introduce learners to reproducible workflows in R using R Markdown. We will discuss what reproducible research is, why it is important, and what common issues hinder reproducibility. The workshop will guide learners through hands-on exercises in R Markdown and show them how to create reproducible reports and share them on GitHub.

SC5 - Introduction to Deep Learning
Short Course
Wed, May 29, 1:30 PM - 5:30 PM
Grand Ballroom G

Instructor(s): Kevin Kuo, RStudio; Javier Luraschi, RStudio

Practical introduction to neural networks with interactive coding exercises in R. We provide an overview of different type of neural network architectures and how they can be applied in a variety of applications.

SC6 - Text Mining with Tidy Data Principles
Short Course
Wed, May 29, 1:30 PM - 5:30 PM
Grand Ballroom I

Instructor(s): Mara Averick, RStudio; Julia Silge, Stack Overflow

Text data is increasingly important in many domains, and tidy data principles and tidy tools can make text mining easier and more effective. In this short course, learn how to manipulate, summarize, and visualize the characteristics of text using these methods and R packages from the tidy tool ecosystem. These tools are highly effective for many analytical questions and allow analysts to integrate natural language processing into effective workflows already in wide use. Explore how to implement approaches such as sentiment analysis of texts, measuring tf-idf, and building text models.

Exhibits Open
SDSS Hours
Wed, May 29, 5:30 PM - 7:00 PM
Grand Ballroom Foyer

PS01 - Opening Mixer & E-Posters
E-Poster
Wed, May 29, 5:30 PM - 7:00 PM
Grand Ballroom Foyer

1
Spatial Statistics and Visualization of Public Health Outcomes
Weichuan Dong, Kent State University
2
Teaching the ASA Guidelines in a Cross-Cultural Setting
Jing Cao, Southern Methodist University
3
The Daily Question: Building Student Trust and Interest in Undergraduate Introductory Probability and Statistics Courses
Matthew A. Hawks, US Naval Academy
4
Extending the Grammar of Graphics beyond ggplot2
Silas Bergen, Winona State University
5
Using Data Science to Support Enrollment Decisions in Higher Education
Monica M King, Drexel University
6
Data-Driven College Admissions: Useful Metrics or Numeric Nonsense?
Emily Rose Flanagan, University of Washington
7
Using Data Verbs to Teach the Management of Tabular Data
Chris John Malone, Winona State University
8
A Shiny Application to Teach the Multiple Linear Regression Analysis in a Undergraduate Course
Carlos M. Lopera-Gómez, Universidad Nacional de Colombia
9
Predicting Matriculation Rates of Dual Enrollment High School Students
Benjamin Kenneth Brown, Oregon Institute of Technology
10
A Meta-analysis on the Effect of Information and Communication Technology Tools in Second Language Acquisition
Songtao Wang, University of Victoria
11
Building Statistical Understanding to Support Organizational Data Culture
Karin Neff, BSD7

SDSS 2019 Hackathon Kickoff
Special Session
Wed, May 29, 6:30 PM - 8:30 PM
Grand Ballroom E

This will be the inaugural year of the Symposium on Data Science and Statistics (SDSS) Hackathon! The goal of the hack is present real world consulting experience that will be mutually beneficial to the industry sponsor and conference participants. Teams will unite participants from diverse academic and industrial backgrounds with statistical and data science skills with the goal of presenting implementable solutions.

We worked in conjunction with the eScience Institute at University of Washington in Seattle to

identify a rich data source and prompt that gives back to the greater Seattle community. Thus, the theme for this year's hackathon will be the housing crisis in the Pacific Northwest that has greatly affected Seattle and Portland. This is a topic that has many perspectives and stakeholders; activists, lawyers, statewide legislature. The datasets we have for the hack present a rich diversity of problems that can be approach from a statistical and data science lens. Participants will be working with data from different levels of geography and from a variety of sources including the American Community Survey, Zillow, Hack Oregon, and other publicly available data pertaining to homelessness and housing insecurity.

This will be a great opportunity for participants to work on a real data problem, learn from professionals in the field, and build relationships with fellow participants, which will enhance the conference experience. We especially encourage students and early career attendees to participate.

Go to the SDSS Events Page to sign-up today!

Thursday, May 30

Exhibits Open
SDSS Hours
Thu, May 30, 7:30 AM - 7:15 PM
Grand Ballroom Foyer

Registration
SDSS Hours
Thu, May 30, 8:00 AM - 6:00 PM
Grand Ballroom Foyer

Speed Mentoring
Special Session
Thu, May 30, 8:00 AM - 9:00 AM
Regency Ballroom AB

Are you looking for a quick way to make connections, solicit career advice, and develop professional relationships? Or maybe you want to provide advice and guidance to early-career statisticians and data scientists? Whether you are interested in mentoring or being mentored, you should consider participating in our new speed mentoring session. Mentees and mentors will have several short, one-on-one, career-focused conversations, followed by unstructured time to socialize and follow up. This is a great opportunity for both mentors and mentees to build their professional networks!

Note: Advance sign-up is required, so please see the SDSS 2019 Events page for details!

GS01 - Welcome and Keynote Address
General Session

Thu, May 30, 9:15 AM - 10:30 AM
Grand Ballroom E

Organizer(s): Kelly McConville, Reed College
Chair(s): Kelly McConville, Reed College

9:30 AM
Generalized Tensor Decompositions for Non-Normal Data
Tamara Kolda, Sandia National Laboratories

CS01 - Teaching Statistics More Effectively to a New Generation of Students
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Grand Ballroom E

Organizer(s): Jo Hardin, Pomona College
Chair(s): Alejandra Castillo, Oregon State University

10:35 AM
Using GitHub with Statistics Undergraduates
Jo Hardin, Pomona College
11:05 AM
Salt Fat Acid Heat: An Alternative to Cookbook Statistics
Andrew Bray, Reed College
11:35 AM
Teaching Data Communication
Amelia McNamara, University of St. Thomas

CS02 - Deciphering Biological Systems via Innovative Statistical Learning Methods
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Grand Ballroom I

Organizer(s): Tian Zheng, Columbia University
Chair(s): Kun Chen, University of Connecticut

10:35 AM
Differential Network Connectivity Analysis
Ali Shojaie, University of Washington
11:05 AM
Modeling Bias in Compositional Data
David Clausen, University of Washington

11:35 AM
Extracting Biological Signals by Controlled Variable Selection
Linxi Liu, Columbia University

CS03 - Open Source and Community
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Grand Ballroom J

Organizer(s): Gabriela de Queiroz, IBM
Chair(s): David Smith, Microsoft

10:35 AM
Getting Involved in Scientific Open Source: Lessons from 7 Years of Growing the ROpenSci Community
Karthik Ram, UC Berkeley
11:05 AM
Sustainers of the Tidyverse
Mara Averick, RStudio
11:35 AM
Building a Community: The R-Ladies Story
Gabriela de Queiroz, IBM

CS04 - Recent Developments in Lower Rank Learning for Complex Data
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Grand Ballroom K

Organizer(s): Xiao-Li Meng, Harvard University
Chair(s): Raymond Wong, Texas A&M University

10:35 AM
MCMC for Dempster-Shafer Statistical Inference
Ruobin Gong, Rutgers University
11:05 AM
Bayesian Analysis of the Covariance Matrix of a Multivariate Normal Distribution with a New Class of Priors
Dongchu Sun, University of Missouri
11:35 AM
Deep Fiducial Inference
Jan Hannig, The University of North Carolina at Chapel Hill

CS05 - Scaling Up Machine Learning to Production
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Regency Ballroom AB

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

10:35 AM
'ML Ops' and Productionizing Machine Learning Workflows
Amy Unruh, Google
11:05 AM
TFX: Production ML Pipelines with TensorFlow
Robert Crowe, Google
11:35 AM
Scalable Automatic Machine Learning with H2O
Erin LeDell, H2O.ai

CS06 - Visual Storytelling
Invited
Thu, May 30, 10:30 AM - 12:05 PM
Regency Ballroom EF

Organizer(s): Silas Bergen, Winona State University
Chair(s): Jerzy Wieczorek, Colby College

10:35 AM
What You Design Is Not What People See
Alberto Cairo, University of Miami
11:05 AM
The Design and Evaluation of Expressive Visualization Tools for Data-Driven Storytelling
Matthew Brehmer, Microsoft Research
11:35 AM
Things We've Learned from Telling the 'Fun' Data Stories
Amber Thomas, The Pudding

CS07 - Reimagining & Introducing New Pedagogy
Contributed
Thu, May 30, 10:30 AM - 12:05 PM
Regency Ballroom C

Chair(s): Julie Zhang, University of Washington

10:35 AM
DATA SCIENCE CERTIFICATION AT MSC – UPR
Abiel Roche-Lima, RCMI-Medical Science School - University of Puerto Rico
10:50 AM
Clinical Data Wrangling: An Active and Didactic Learning Workshop
Ted Laderas, Oregon Health & Science University
11:05 AM
What Can Data Science Look Like in High School?
Tim Erickson, Epsitemological Engineering and Lick-Wilmerding High School
11:20 AM
Teaching Upper Level Statistics Courses through a Shared/Hybrid Model
Jingchen Hu, Vassar College
11:35 AM
Data Science and the Pedagogical Reform of Introductory Statistics
Brendan Patrick Purdy, Moorpark College
11:50 AM
Floor Discussion

CS08 - SADM Invited Papers
Invited
Thu, May 30, 1:30 PM - 3:05 PM
Grand Ballroom I

Organizer(s): Bertrand Clarke, University of Nebraska-Lincoln; Jia Li, Penn State University
Chair(s): Aaron Molstad, Fred Hutchinson Cancer Research Center

1:35 PM
Bayesian Variable Selection in High-Dimensional EEG Data Using Spatial Structured Spike and Slab Prior
Dipak K. Dey, University of Connecticut
2:05 PM
Mean Residual Function: a Tool for Exploring Patterns in Big Data
Ehsan S. Soofi, University of Wisconsin-Milwaukee
2:35 PM
Slow-kill for Big Data Learning
Yiyuan She, Florida State University

CS09 - Project Jupyter
Invited
Thu, May 30, 1:30 PM - 3:05 PM

Regency Ballroom AB

Organizer(s): Brian Granger, Cal Poly; Fernando Perez, UC Berkeley
Chair(s): Casey Jelsema, West Virginia University

1:35 PM
Sharing Reproducible Computations on Binder
Lindsey J. Heagy, UC Berkeley
2:05 PM
Open Infrastructure in the Cloud with JupyterHub
Chris Holdgraf, UC Berkeley
2:35 PM
JupyterLab: An Extensible and Flexible Platform for Collaborative Data Science
Brian Ellison Granger, Cal Poly / Project Jupyter

CS10 - Data Science's X-Factor
Invited
Thu, May 30, 1:30 PM - 3:05 PM
Regency Ballroom C

Organizer(s): Katherine M. Kinnaird, Smith College
Chair(s): Mine Dogucu, .

1:35 PM
Student Difficulties in Data Science Instruction: Early Findings
Karl R. B. Schmitt, Valparaiso University
2:05 PM
Data Science In/Among/With/Toward the Humanities
John Laudun, University of Louisiana
2:35 PM
Data Physicalizations: Where Art, Data, and Domain Applications Combine
Katherine M. Kinnaird, Smith College

CS11 - Data Visualization in Python
Invited
Thu, May 30, 1:30 PM - 3:05 PM
Regency Ballroom EF

Organizer(s): Todd Iverson, Winona State University
Chair(s): Todd Iverson, Winona State University

1:35 PM

Introduction to Visualization with Python
Stephen F. Elston, Quantia Analytics, LLC
2:05 PM
Altair: Declarative Visualization in Python - Part 1
Dominik Moritz, University of Washington
2:35 PM
Altair: Declarative Visualization in Python - Part 2
Kanit "Ham" Wongsuphasawat, Apple

CS12 - Enterprise Applications of Data Science
Contributed
Thu, May 30, 1:30 PM - 3:05 PM
Grand Ballroom J

Chair(s): Gabriela de Queiroz, IBM

1:35 PM
Estimating Causal Effects in Large Scale Online Experiments and Designing Automated A/B Testing Platforms for Machine Learning
Zuzanna Klyszejko, MongoDB
1:50 PM
Data Storytelling: Improve Insight-To-Action Conversion for a Greater Real World Impact
Yu Zhou, Mastercard
2:05 PM
Detecting Innovative Companies via Their Website
Piet Daas, Statistics Netherlands
2:20 PM
Metrics and Modeling in Large-Scale Digital Experimentation
W. Duncan Wadsworth, Microsoft
2:35 PM
Forecasting at Scale to Champion Customer Trust
Ana Bertran, Salesforce
2:50 PM
Floor Discussion

CS13 - Computationally Intensive Methods: Resampling and MCMC
Contributed
Thu, May 30, 1:30 PM - 3:05 PM
Grand Ballroom K

Chair(s): Honglang Wang, Indiana University-Purdue University Indianapolis

1:35 PM

Jackknife Empirical Likelihood Approach for K-Sample Tests via Energy Distance
Yongli Sang, University of Louisiana at Lafayette
1:50 PM
Gelman-Rubin: Improved Stability and a Principled Threshold
Christina Phan Knudson, University of St. Thomas
2:05 PM
Error Estimation for Randomized Numerical Linear Algebra via the Bootstrap
Miles Lopes, UC Davis
2:20 PM
A Scalable Regression Estimation Procedure for Competing Risks Data
Eric S. Kawaguchi, University of California, Los Angeles
2:35 PM
Floor Discussion

PS02 - Data Science Applications E-Posters, I
E-Poster
Thu, May 30, 3:00 PM - 4:00 PM
Grand Ballroom Foyer

1
Automated Survey Text Analysis -- Supervised Latent Dirichlet Allocation (SLDA)
Christine P. Chai, Microsoft
2
Comparing various string similarity algorithms in the task of name-matching
Aleksandra Zaba, University of Utah
3
Hypothesis Testing in Nonlinear Function on Scalar Regression with Application to Child Growth Study
Mityl Biswas, NC State University
4
Comparing Object Correlation Metrics for Effective Space Traffic Management
Julie Zhang, University of Washington
5
Batch effect adjustment via ensemble learning in the validation of genomic classifiers
Yuqing Zhang, Boston University
6
Tensor Mixed Effects Model with Application to Nanomanufacturing Inspection
Xiaowei Yue, Virginia Polytechnic Institute and State University
7
Burst Detection in Call Trains for Identifying Fraud in Telecommunications
Miguel Raul Pebes Trujillo, Indiana University Bloomington, Department of Statistics
8
Active Labeling using Model-based Classification

Min Fang, San Jose State University
9
Analyzing Influence of Social Media Through Twitter
Dhrubajyoti Ghosh, North Carolina State University
10
Diversity of forest structure across the United States
Jessica Lynn Gilbert, Purdue University
11
ClusterJob, an Experiment Management System For Ambitious Data Science
Bekk Blando, Clemson University
12
A Maximum Likelihood Method for Correlated Discrete and Continuous Outcomes with Selection, Lagged Effects and Variance
Rhoda Nandai Muse, University of Arizona, Mathematics Department
13
Gender Distribution in Movie Roles
Vijay Ravuri, CalPoly SLO
14
Evaluating and forecasting the CD4 cell count evolution in HIV+ patients from a Bayesian stochastic model related to the logistic curve with multiple inflection points.
Victor Cruz-Torres, University of Puerto Rico

CS14 - The IMS Program on Probabilistic Views of Machine Learning
Invited
Thu, May 30, 4:00 PM - 5:35 PM
Grand Ballroom I

Organizer(s): Eric Chi, North Carolina State University; Brad Price, West Virginia University
Chair(s): Brad Price, West Virginia University

4:05 PM
Prediction with Confidence – General Framework for Predictive Inference
Regina Liu, Rutgers University
4:35 PM
Scalable and Model-free Methods for Multiclass Probability Estimation
Helen Zhang, University of Arizona
5:05 PM
Fiducial Made Sexy: Statistical Inference for Machine Learning Problems
Thomas Lee, UC Davis

CS15 - Linguistic Diversity in NLP
Invited
Thu, May 30, 4:00 PM - 5:35 PM
Grand Ballroom J

Organizer(s): Rachael Tatman, Kaggle
Chair(s): Julia Silge, Stack Overflow

4:05 PM
An Introduction to Computational Sociolinguistics
Rachael Tatman, Kaggle
4:35 PM
English Isn't Generic for Language, Despite What NLP Papers Might Lead You to Believe
Emily M. Bender, University of Washington
5:05 PM
Learning the Language of BlackTwitter
Brandeis Hill Marshall, Spelman College

CS16 - Recent Advances in Matrix and Tensor Factorization Models
Invited
Thu, May 30, 4:00 PM - 5:35 PM
Grand Ballroom K

Organizer(s): Raymond Wong, Texas A&M University
Chair(s): Jan Hannig, The University of North Carolina at Chapel Hill

4:35 PM
Linked Matrix Factorization
Eric F. Lock, University of Minnesota
5:05 PM
Boosted Sparse and Low-Rank Tensor Regression
Kun Chen, University of Connecticut

CS17 - Shared Infrastructure for Data Science
Invited
Thu, May 30, 4:00 PM - 5:35 PM
Regency Ballroom AB

Organizer(s): Soren Harner, Permaling
Chair(s): James Sharpnack, UC Davis

4:05 PM
The Machine Learning Lifecycle with MLflow
Siddharth Murching, Databricks, Inc.
4:35 PM
Low-Latency Model Serving with MLflow and MLeap

Corey Zumar, Databricks, Inc.
5:05 PM
Bayesian Structured Time Series in TensorFlow Probability
Jacob Burnim, Google

CS18 - Communication Within and Beyond the Modern Data Science/Statistics Classroom

Invited
Thu, May 30, 4:00 PM - 5:35 PM
Regency Ballroom C

Organizer(s): Alicia Johnson, Macalester College
Chair(s): Christina Phan Knudson, University of St. Thomas

4:05 PM
Agile, Reproducible, and Accessible: Using Bookdown for Communication Within and Beyond the Classroom
Alicia Johnson, Macalester College
4:35 PM
Using Slack for Communication and Collaboration in the Classroom
Albert Y. Kim, Smith College
5:05 PM
Using Blogdown to Connect Beyond the Classroom
Alison Hill, RStudio

CS19 - Statistical Modeling in Python

Invited
Thu, May 30, 4:00 PM - 5:35 PM
Regency Ballroom EF

Organizer(s): Dennis Sun, Cal Poly
Chair(s): Kelly Nicole Bodwin, Cal Poly - San Luis Obispo

4:05 PM
Linear Modeling in Python with SALMON
Alex Boyd, University of California, Irvine
4:35 PM
A Grammar of Data Analysis
Dennis Sun, Google
5:05 PM
Symbulate: Probability Simulations in Python
Kevin Ross, Cal Poly

PS03 - Data Science Applications E-Posters, II
E-Poster
Thu, May 30, 5:30 PM - 6:30 PM
Grand Ballroom Foyer

1
Automated Analytics of the Solar Corona with Scalable Cloud Based Platforms
Lars K. S. Daldorff, JHU/APL
2
Modeling and Forecasting the Percent Changes in the National Park Visitation Counts Using Social Media Data
Russell Goebel, Western Washington University
3
Estimating Plant Growth Curves and Derivatives by Modeling Crowdsourced Imaged-Based Data
Haozhe Zhang, Iowa State University
4
Using Bayesian Networks to Perform Reject Inference
Billie Anderson, Harrisburg University
5
Usability evaluation of data presentation for official statistics
Lin Wang, U.S. Census Bureau
6
Do Unregistered Voters Want to Vote? Automatic Registration and Oregon Elections Turnout.
Matthew Stephan Yancheff, Reed College
7
Relationship between physical activity and depression in elderly Costa Ricans
Shu Li, Kent State University
8
Building an Interpretable Incident Prediction model for Site Reliability
Jiaping Zhang, Salesforce
9
For-estimation: Post-stratification to increase efficiency of forest attribute estimates
Miranda Rintoul, Reed College
10
Forecasting NBA Fan Support using Time Series Analysis
Victor Wilson, Cal Poly San Luis Obispo
11
Handling Missing Data in Cardiovascular Disease Prediction Using Neural Networks
Megan Shand, Broad Institute
12
Leverage Machine Learning to Advance Risk Prediction with Electronic Health Record
Yirui Hu, Geisinger
13

Multiple uses for chronic condition data mart
John Massman, Virginia Mason
14
Team Item Response Models
Deborshee Sen, Duke University

GS02 - Symposium on Data Science and Statistics Banquet
General Session
Thu, May 30, 6:30 PM - 8:00 PM
Grand Ballroom E

Organizer(s): Kelly McConville, Reed College
Chair(s): Jennifer L. Beaumont, Terasaki Research Institute

7:00 PM
Statistics Isn't All That Funny, but it Has Its Moments
Joel Grus, Allen Institute for Artificial Intelligence
Friday, May 31

Exhibits Open
SDSS Hours
Fri, May 31, 7:30 AM - 3:45 PM
Grand Ballroom Foyer

Registration
SDSS Hours
Fri, May 31, 7:30 AM - 5:30 PM
Grand Ballroom Foyer

GS03 - Friday Keynote Address
General Session
Fri, May 31, 8:30 AM - 9:45 AM
Grand Ballroom E

Organizer(s): Kelly McConville, Reed College
Chair(s): Jo Hardin, Pomona College

8:35 AM
Data Science: How the Union of Inferential Thinking and Computation Are Transforming Research and Education at Berkeley
Fernando Perez, UC Berkeley
9:35 AM
Sponsor Spotlight - SAS

9:40 AM
Floor Discussion

PS04 - Machine Learning E-Posters, I
E-Poster
Fri, May 31, 9:45 AM - 10:45 AM
Grand Ballroom Foyer

2
Artificial Intelligence Mammography Model and Healthcare Savings Opportunity
Olajide Israel Ajayi, Blue Cross NC
3
The Geometry of feature embeddings in kernel discriminant analysis-deterministic or randomized
Jiae Kim, The Ohio State University
4
HARNESSING the POWER of MACHINE LEARNING METHODS in HIV VIROLOGIC FAILURE RISK PREDICTION
Allan Kimaina, brown university
5
Practical Considerations of Deep Learning in Digital Pathology
Shubing Wang, Merck
6
Identifying Shifts in Forest Communities Using Machine Learning Techniques
Trenton W Ford, University of Notre Dame
7
Rapid deployment of a Machine Learning-based derived biomarker using publicly available data sources for covariate adjusted descriptive modeling.
Albert Taylor, Origent Data Sciences
8
Adaptively Stacked Ensembles for Influenza Forecasting with Incomplete Data
Thomas Charles McAndrew, University of Massachusetts Amherst
9
Overcoming Big Data: Linking the 2014 National Hospital Care Survey to the 2014/2015 Medicare CMS Master Beneficiary Summary File
Scott Robert Campbell, National Opinion Research Center at University of Chicago
10
Comparing Performance of Lasso, Group Lasso, and Linear Regression with Categorical Predictors
Yihuan Huang, UCLA
12
ML-assisted ongoing monitoring for fighting fraud and abuse
Jose Ferreira, Google
13
Time-aggregated forecasting for ultra high dimensional regression and time-series error
Sayar Karmakar, University of Florida
14
Empirical priors for prediction in sparse high-dimensional linear regression

Yiqi Tang, NC State University

CS20 - Data Science Platforms: Spark
Invited
Fri, May 31, 10:30 AM - 12:05 PM
Grand Ballroom E

Organizer(s): Kevin Kuo, RStudio
Chair(s): Kevin Kuo, RStudio

10:35 AM
An R Interface to Hail
Michael Lawrence, Genentech Research
11:05 AM
Scaling Sparklyr with Streams and Arrow
Javier Luraschi, RStudio
11:35 AM
Interpretable Machine Learning Using rsparkling
Navdeep Gill, H2O.ai

CS21 - A Field Guide to Education Tools in Data Science
Invited
Fri, May 31, 10:30 AM - 12:05 PM
Grand Ballroom I

Organizer(s): Alison Hill, RStudio
Chair(s): Alison Hill, RStudio

10:35 AM
Necessity Is the Mother of Invention: Evolution of a Data Science Team
Adrienne Zell, Oregon Health and Science University
11:05 AM
Using Unit Testing to Teach Data Science
Kyle Gorman, CUNY
11:35 AM
Data Presentation For Everyone: Simple Ways to Educate without Teaching
Allison Sliter, Digimarc Inc

CS22 - Building and Growing Data Science Teams
Invited

Fri, May 31, 10:30 AM - 12:05 PM
Grand Ballroom J

Organizer(s): Jacqueline Nolis, Nolis, LLC
Chair(s): Jacqueline Nolis, Nolis, LLC

10:35 AM
From Zero to A^X: Scaling Data Science Teams
Amanda Casari, Google Cloud
11:05 AM
Together at Last: Heterogeneous Teams and the Key to Success
Heather Nolis, T-Mobile
11:35 AM
Creating Effective Data Science Teams
Mehar Singh, ProCogia

CS23 - Advances in Analysis and Computing in Complex Data
Invited
Fri, May 31, 10:30 AM - 12:05 PM
Grand Ballroom K

Organizer(s): George Michailidis, University of Florida
Chair(s): Regina Liu, Rutgers University

10:35 AM
Graph-Based Change-Point Detection
Lynna Chu, UC Davis
11:05 AM
A Double Core Tensor Factorization and Its Applications to Heterogeneous Data
George Michailidis, University of Florida
11:35 AM
Individualized Fusion Learning (IFusion) with Applications to Personalized Inference
Minge Xie, Rutgers University

CS24 - Recent Developments on Machine Learning
Invited
Fri, May 31, 10:30 AM - 12:05 PM
Regency Ballroom AB

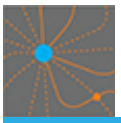
Organizer(s): Xiaotong Shen, University of Minnesota
Chair(s): Xiaotong Shen, University of Minnesota

10:35 AM
Shrinking Characteristics of Precision Matrix Estimators
Adam J. Rothman, University of Minnesota
11:05 AM
P-Splines with an L1 Penalty for Repeated Measures
Hui Jiang, University of Michigan
11:35 AM
Community Detection with Dependent Connectivity
Annie Qu, University Illinois at Urbana-Champaign

CS25 - Software Packages for Data Science
Contributed
Fri, May 31, 10:30 AM - 12:05 PM
Regency Ballroom C

Chair(s): Amrina Ferdous, Boise State University

10:35 AM
An R Package for Linear Mediation Analysis with Complex Survey Data
Yujiao Mai, St. Jude Children's Research Hospital
10:50 AM
GREIN: An Interactive Web Platform for Re-Analyzing GEO RNA-Seq Data
Naim Al Mahi, University of Cincinnati
11:05 AM
Bioc2mlr: R Package to Bridge Between Bioconductor's S4 Complex Genomic Data Container, to Mlr, a Meta Machine Learning Aggregator Package.
Dror Berel, Fred Hutch



CS26 - Data Visualization in Applications
Contributed
Fri, May 31, 10:30 AM - 12:05 PM
Regency Ballroom EF

Chair(s): Oyeleke Olaoye, .

10:35 AM
Topological Data Analysis for Understanding Phenotypic Presentation in Aortic Stenosis
Sirish Shrestha, West Virginia University
10:50 AM
Assessing and Visualizing the Impact of Medical Coding Systems for Predicting Inpatient Mortality
Brian Hochrein, IBM Watson Health
11:05 AM
Methods for Visualizing Dimension Reduction in R

Tiffany Jiang, UC Davis
11:20 AM
Floor Discussion

CS27 - Data Science Platforms: Deep Learning
Invited
Fri, May 31, 1:30 PM - 3:05 PM
Grand Ballroom E

Organizer(s): Javier Luraschi, RStudio
Chair(s): Javier Luraschi, RStudio

1:35 PM
Deep Learning and Probabilistic Programming with Applications to Intelligent Reality
Soren Harner, Permaling
2:05 PM
R Interfaces to TensorFlow and Keras
Kevin Kuo, RStudio
2:35 PM
Deep Learning Models at Scale with Apache Spark
Joseph Kurata Bradley, Databricks, Inc.

CS28 - Data Science Ethics Meet Reality
Invited
Fri, May 31, 1:30 PM - 3:05 PM
Grand Ballroom J

Organizer(s): Os Keyes, University of Washington
Chair(s): Brandeis Hill Marshall, Spelman College

1:35 PM
The Politics of Data
Meg Drouhard, University of Washington
2:05 PM
The Political Consequences of Repurposing Data
Meg Young, University of Washington
2:35 PM
Beyond Methodological Rigor: Widening the Scope of Ethics in Data Science
Anissa Tanweer, University of Washington

CS29 - The Cutting Edge in Statistical Machine Learning
Invited
Fri, May 31, 1:30 PM - 3:05 PM
Regency Ballroom AB

Organizer(s): Daniela Witten, University of Washington
Chair(s): Boxiang Wang, University of Iowa

1:35 PM
A Continuous-Time View of Early Stopping in Least Squares Regression
Ryan Tibshirani, Carnegie Mellon University
2:05 PM
Fused Lasso on Graphs: Applications to Nonparametric Statistical Problems
Oscar Hernan Madrid Padilla, UC Berkeley
2:35 PM
Two-Stage Computational Framework for Sparse Generalized Eigenvalue Problem
Kean Ming Tan, University of Minnesota

CS30 - Data Visualization Education
Invited
Fri, May 31, 1:30 PM - 3:05 PM
Regency Ballroom EF

Organizer(s): Silas Bergen, Winona State University; Amelia McNamara, University of St. Thomas
Chair(s): Silas Bergen, Winona State University

1:35 PM
Teaching Data Visualization: Integrating Theory and Practice
Michael Freeman, University of Washington
2:05 PM
A Three-Part Data Visualization Curriculum
Jerzy Wieczorek, Colby College
2:35 PM
Help Me Understand: Guiding Visualization Users with Annotations
Robert Kosara, Tableau Software

CS31 - Instructional Applications & Insights
Contributed
Fri, May 31, 1:30 PM - 3:05 PM
Grand Ballroom I

Chair(s): Emily Rose Flanagan, University of Washington

1:35 PM
Apply “STEAMS” Methodology on Managing Europe Travel
Charles Chen, Applied Materials
1:50 PM
A Robust and Dynamic Formulation for Predicting Student Offer Acceptance
Michael Liut, McMaster University
2:05 PM
P-Values: A Closer Look
Jeanne Li, Santa Barbara Cottage Hospital
2:20 PM
Floor Discussion

CS32 - Statistical Methods for Analyzing Large Scale or Massive Data
Contributed
Fri, May 31, 1:30 PM - 3:05 PM
Grand Ballroom K

Chair(s): Alona Kryshchenko, California State University Cannel Islands

1:35 PM
High-Dimensional Association Detection in Large Scale Genomic Studies
Hillary Koch, Pennsylvania State University
1:50 PM
Threshold Knot Selection for Large-Scale Spatial Models with Applications to the Deepwater Horizon Disaster
Casey Jelsema, West Virginia University
2:05 PM
Goodness-of-Fit Tests for Large Data Sets
Taras Lazariv, TU Dresden
2:20 PM
Big Data and Portfolio Optimization
QIYU WANG, Zhejiang Univ of Finance and Econ
2:35 PM
An Application of Linear Programming to Computational Statistics
John M. Ennis, Aigora
2:50 PM
Accelerate Pseudo-Proximal Map Algorithm and Its Application to Network Analysis
Dao Nguyen, University of Mississippi

Hackathon Update
Special Session
Fri, May 31, 1:30 PM - 3:05 PM
Regency Ballroom C

Join the Hackathon participants as they present their findings.

PS05 - Machine Learning E-Posters, II
E-Poster
Fri, May 31, 3:00 PM - 4:00 PM
Grand Ballroom Foyer

1
Clustering Chocolate Types: Dark, White, Milk and Fruit
Kaitlyn Zhang, Stanford OHS
2
Statistical Approaches for Identifying Untargeted Metabolites Prognostic for Kidney Disease Progression in Type 2 Diabetic Patients: Application to the Chronic Renal Insufficiency Cohort Study
Jing Zhang, UCSD Moores Cancer Center
3
Genomic Determination Index
Cheng Cheng, St. Jude Children's Research Hospital
4
On Combining Data from Distinct Nonlinear Predictive Models
Amrina Ferdous, Boise State University
5
Predicting Unknown Links for Interconnected Networks
Yubai Yuan, UIUC
6
A Bayesian Structural Time Series-Based Approach for Understanding and Predicting Temperatures in the Red Sea
Nabila Bounceur, King Abdullah University of Science and Technology
7
Is robustness trade-off really inevitable?
Jungeum Kim, Purdue Department of Statistics
8
HARNESSING THE POWER OF MACHINE LEARNING METHODS IN PROSPECTIVE HIV CARE AND TREATMENT
Allan Kimaina, brown university
9
Machine Learning meets Survival Analysis for the personalized medicine
Jongyun Jung, University of Nevada, Las Vegas
10
Predicting Claims Litigation using Text Mining
Xiyue Liao, Universiry of California, Santa Barbara
11
A Multicategory Kernel Distance Weighted Discrimination Method for Multiclass Classification
Boxiang Wang, University of Iowa
13
Comparison of Automated Liver Image Quality Evaluation Using Handcrafted Features and

Convolutional Neural Networks
Wenyi Lin, University of California, San Diego
14
Statistical Learning on Next-Generation Sequencing of T cell Repertoire Data
Li Zhang, UCSF

CS33 - Backend Data Science
Invited
Fri, May 31, 3:40 PM - 5:15 PM
Grand Ballroom E

Organizer(s): Edgar Ruiz, RStudio
Chair(s): Soren Harner, Permaling

3:45 PM
Data Science with Databases and R
James Blair, RStudio
4:15 PM
STOIC Next-Generation Spreadsheet: Bringing Data Science to the Masses
Ismael Ghalimi, STOIC
4:45 PM
Working with Images and Text in R Through Embeddings
Michael Lucy, Basilica

CS34 - Computational Statistics for Large-Scale Biological Data
Invited
Fri, May 31, 3:40 PM - 5:15 PM
Grand Ballroom K

Organizer(s): Jacob Bien, University of Southern California
Chair(s): Kean Ming Tan, University of Minnesota

3:45 PM
Computationally Efficient High-Dimensional Interaction Modeling
Guo Yu, University of Washington
4:15 PM
Inference for Diversity Under Networked Models
Bryan Martin, University of Washington
4:45 PM
Variance Component Testing and Selection for a Longitudinal Microbiome Study
Jin Zhou, University of Arizona

CS35 - Modern Multivariate Analysis
Invited
Fri, May 31, 3:40 PM - 5:15 PM
Regency Ballroom AB

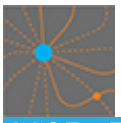
Organizer(s): Adam J. Rothman, University of Minnesota
Chair(s): Adam J. Rothman, University of Minnesota

3:45 PM
The Multivariate Square Root Lasso: Computational and Theoretical Insights
Aaron Molstad, Fred Hutchinson Cancer Research Center
4:15 PM
Estimating Multiple Precision Matrices Using Cluster Fusion Regularization
Brad Price, West Virginia University
4:45 PM
L_2-Regularization and Some Path-Following Algorithms
Yunzhang Zhu, The Ohio State University

CS36 - Democratizing Data Science with Workflows
Invited
Fri, May 31, 3:40 PM - 5:15 PM
Regency Ballroom C

Organizer(s): Michael I. Love, UNC-Chapel Hill
Chair(s): Stas Kolenikov, Abt Associates

3:45 PM
Publishing Literate Programming Workflows in Scientific Journals
Michael I. Love, UNC-Chapel Hill
4:15 PM
When Should You Add Github, Make and Docker to Your Data Science Workflow?
Tiffany Timbers, University of British Columbia
4:45 PM
Useful Tools for Teaching and Outreach in Data Science: Workflows, Case Studies, Github Classroom, and Slack
Stephanie Hicks, Johns Hopkins Bloomberg School of Public Health



CS37 - Data Visualizations at the Institute for Health Metrics and Evaluation
Invited
Fri, May 31, 3:40 PM - 5:15 PM

Regency Ballroom EF

Organizer(s): Brian Dart, IHME
Chair(s): Disha Patel, University of Washington

3:45 PM
Building Interactive Data Visualization for a Global (Health) Audience
Ryan Shackleton, University of Washington
4:15 PM
The Story of a Chart: Data Visualization Principles to Simplify Complexity
Evan Laurie, University of Washington
4:45 PM
Behind the Scenes: Building Tools to Visualize Intermediate Results in Complex Data Science Pipelines
Marlena Bannick, University of Washington

CS38 - Engaging Students in Statistics & Data Science

Contributed
Fri, May 31, 3:40 PM - 5:15 PM
Grand Ballroom I

Chair(s): Ted Laderas, Oregon Health & Science University

3:45 PM
STEAMS Approach on Playing Video Games
Mason Chen, Stanford OHS
4:00 PM
Competition Based Teaching of Machine Learning
Mikael Vejdemo-Johansson, CUNY College of Staten Island
4:15 PM
USING R and SPSS for TEACHING STATISTICS
Lucy Xiaojing Kerns, Youngstown State University
4:30 PM
Tools for R in Introductory Statistics Courses
Kelly Nicole Bodwin, Cal Poly - San Luis Obispo
4:45 PM
Teaching Data Science Students to Write Clean Code
Todd Iverson, Winona State University
5:00 PM
Hack Weeks as a Model for Data Science Education and Collaboration
Daniela Huppenkothen, University of Washington

CS39 - Data and Society

Contributed
Fri, May 31, 3:40 PM - 5:15 PM
Grand Ballroom J

Chair(s): Heather Nolis, T-Mobile

3:45 PM
Using Convolutional Neural Networks to Automatically Classify Logos on Shopping Receipts
Émilie Mayer, Statistics Canada
4:00 PM
Using Topological Data Analysis to Assess Gerrymandering in Voting Districts
Courtney Thatcher, University of Puget Sound
4:15 PM
Predicting the Success of an Crowdfunding Campaign: Spatial Location-Based Trajectory Modeling
Han Yu, University of Northern Colorado
4:30 PM
Nurturing select customers using a state-space model (Investment Recommender / Resource allocation)
Eunice Kim, Microsoft
4:45 PM
Floor Discussion

CS40 - SAS Open-Source Platforms for Analytics
Invited
Fri, May 31, 5:20 PM - 6:25 PM
Grand Ballroom E

Organizer(s): Jim Harner, West Virginia University
Chair(s): Wendy Martinez, Bureau of Labor Statistics

5:25 PM
SAS Viya: A Modern Scalable and Open Platform for Artificial Intelligence
Wayne Thompson, SAS
5:55 PM
Making Predictive Modeling Approachable with JMP Pro
Jordan Hiller, JMP

CS41 - Incorporating Ethics and Inclusion in Undergraduate Statistics Curriculum
Invited
Fri, May 31, 5:20 PM - 6:25 PM
Grand Ballroom I

Organizer(s): Brianna Heggeseth, Macalester College

Chair(s): Jingchen Hu, Vassar College

5:25 PM
Ethics in an Advanced Undergraduate Seminar: Statistical Analysis of Social Network Data
Miles Q. Ott, Smith College
5:55 PM
Intertwining Data Ethics into Intro Stats
Brianna Heggeseth, Macalester College

CS42 - Interoperability: Your R Package Can Depend on Its Friends
Invited
Fri, May 31, 5:20 PM - 6:25 PM
Regency Ballroom C

Organizer(s): Matthew N. McCall, University of Rochester
Chair(s): Xiaowei Yue, Virginia Polytechnic Institute and State University

5:25 PM
Case Studies in Interoperability: From Generic Classes to Specific Functions
Matthew N. McCall, University of Rochester
5:55 PM
How Core Data Structures Drive Interoperability in the Bioconductor Project
Marcel Ramos, CUNY SPH

CS43 - Grammar of Graphics: The Twentieth Anniversary
Invited
Fri, May 31, 5:20 PM - 6:25 PM
Regency Ballroom EF

Organizer(s): Jim Harner, West Virginia University
Chair(s): Claus Wilke, University of Texas at Austin

5:25 PM
Past, Present, and Future of Grammar of Graphics Systems
Lee Wilkinson, H2O.ai
5:55 PM
Discussant
Anushka Anand, Tableau
6:05 PM
Discussant
Jeffrey Heer, University of Washington
6:15 PM

Discussant
Bryan Van de Ven, Microsoft

CS44 - Science and the Environment
Contributed
Fri, May 31, 5:20 PM - 6:25 PM
Grand Ballroom J

Chair(s): Melanie Edwards, Exponent, Inc.

5:25 PM
Trend Assessment for Daily Snow Depths with Changepoints Considerations
Jaechoul Lee, Boise State University
5:40 PM
Yield Forecasting Based on Short Time Series with High Spatial Resolution Data
Yuzhen Zhou, University of Nebraska Lincoln
5:55 PM
Are Forest Communities Impacted by Climate Change?
Jonathan Andrew Knott, Purdue University
6:10 PM
Extracting Signal from the Noisy Environment of an Ecosystem
Pranita Pramod Patil, Harrisburg University of Science & Technology

CS45 - Change Point Detection
Contributed
Fri, May 31, 5:20 PM - 6:25 PM
Grand Ballroom K

Chair(s): Dao Nguyen, University of Mississippi

5:25 PM
Detection of Structural Changes in Correctly Specified and Misspecified Conditional Quantile Polynomial Distributed Lag (QPDL) Model Using Change-Point Analysis
KWADWO AGYEI NYANTAKYI, GHANA INSTITUTE OF MANAGEMENT AND PUBLIC ADMINISTRATION
5:40 PM
Robust Graph Change-Point Detection for Brain Evolvement Study
Honglang Wang, Indiana University-Purdue University Indianapolis
5:55 PM
Graph Theoretic Statistics for Change Detection and Localization in Multivariate Data
Matthew A. Hawks, US Naval Academy
6:10 PM
Floor Discussion

CS46 - Recent Advancements in Deep Learning
Contributed
Fri, May 31, 5:20 PM - 6:25 PM
Regency Ballroom AB

Chair(s): Yunzhang Zhu, The Ohio State University

5:25 PM
Statistical Evaluation of Long Memory in Recurrent Neural Networks
Alexander Greaves-Tunnell, University of Washington
5:40 PM
On Interpretable Machine Learning
Serge Berger, Microsoft
5:55 PM
Machine Learning Methods for Modeling Animal Movement
Dhanushi Wijeyakulasuriya, Pennsylvania State University
6:10 PM
Optimal Transport Classifier: Defending Against Adversarial Attacks by Regularized Deep Embedding
Yao Li, University of California, Davis
Saturday, June 1

Registration
SDSS Hours
Sat, Jun 1, 7:30 AM - 2:00 PM
Grand Ballroom Foyer

GS04 - Fireside Chat
General Session
Sat, Jun 1, 8:30 AM - 9:30 AM
Grand Ballroom E

Chair(s): Gabriela de Queiroz, IBM

New horizons and controversies seem to emerge constantly in the world of statistics and data science. Who can keep up? Our distinguished panel of statistics and data science leaders will discuss this and more in an informal and wide-reaching conversation that contextualizes the SDSS experience with issues of the day.

8:35 AM
Fireside Chat Panel
Amanda Casari, Google Cloud; Amelia McNamara, University of St. Thomas; Mara Averick, RStudio; Miguel Marino, OHSU-PSU School of Public Health

PS06 - Computational Statistics E-Posters
E-Poster
Sat, Jun 1, 9:30 AM - 10:30 AM
Grand Ballroom Foyer

1
Application of Dynamic Bi-Partite Stochastic Block Models
Neil Hwang, CUNY-Bronx Community College
2
Estimation of Semiparametric Functional Coefficients Panel Data Model
Shaymal C Halder, Auburn University
3
Discovery of Gene Regulatory Networks Using Adaptively Selected Gene Perturbation Experiments
Michele Zemplenyi, Harvard University
4
A Computational Approach to the Structure of Subtraction Games
Kali Lacy, Purdue University
5
Covariate Information Number for Feature Screening in Ultrahigh-Dimensional Supervised Problems
Debmalya Nandy, Penn State University
6
A Data-Adaptive Targeted Learning Approach of Evaluating Viscoelastic Assay Driven Trauma Treatment Protocols
Linqing Wei, UC Berkeley, Department of Biostatistics
7
Approximate Fiducial Computation and Deep Fiducial Inference
Gang Li, The University of North Carolina at Chapel Hill
8
Innovative Robust Boosting Algorithms
Zhu Wang, UT Health San Antonio
9
A Model Based Data Fusion Algorithm using Bayesian Hierarchal Modeling for Density Estimation of Rare Species
Purna Gamage, Wake Forest University
10
Kernel-estimated Nonparametric Overlap-Based Syncytial Clustering
Israel A Almodovar-Rivera, University of Puerto Rico-Medical Science Campus
11
Developing Nonlinear Genetic Signatures for Enzalutamide Resistance in Prostate Cancer
Isaac Zhao, Brown University
12
Approximate Bayesian Computational Statistical Methods to Estimate the Strength of Divergent Selection in Yeast
Martyna Lukaszewicz, University of Idaho

13
Wavelet Shrinkage Using Bayesian False Discovery Rate Methods: a Comparison Study
Rodney Vasconcelos Fonseca, Unicamp
14
Analyzing Air Traffic Data with Spark-GraphX
Chathurangi Heshani Pathiravasan, Southern Illinois University, Carbondale

SDSS Teaching Data Science Workshop for High School Teachers
Special Session
Sat, Jun 1, 9:30 AM - 11:30 AM
Regency Ballroom EF

Instructor(s): Shannon Ellis, UC San Diego

Considering how to incorporate data science into your high school STEM classroom?

The goal of this workshop is for you to leave with data science skills and applicable examples that can be used in your classroom.

This workshop will answer questions like:

• What is data science?

• How can high schoolers prepare for data science courses in college?

• What does a career in data science involve?

We will walk through how data scientists carry out projects using RStudio, introduce the basics of the R programming language, and work with real datasets to generate visualizations and analyze data.

Note: Advance sign-up is required, so please see the SDSS 2019 Events page for details!

CS47 - Data Science for Fun
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM
Grand Ballroom E

Organizer(s): David Smith, Microsoft
Chair(s): Ana Bertran, Salesforce

10:05 AM
Minecraft, R, and Containers
David Smith, Microsoft
10:35 AM
Using Deep Learning in R to Generate Offensive License Plates
Jacqueline Nolis, Nolis, LLC

CS48 - Recent Advances in Statistical Network Analysis
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM
Grand Ballroom I

Organizer(s): James L Rosenberger, NISS; Lingzhou Xue, Penn State University and NISS
Chair(s): Hyun Bin Kang, Western Michigan University

10:05 AM
Statistical estimation of network models from egocentrically sampled network data
Jeanette Kurian Birnbaum, University of Washington
10:35 AM
Model-based clustering of large networks
David Hunter, Penn State University
11:05 AM
Temporal Exponential-Family Random Graph Models with Time-Evolving Latent Block Structure for Dynamic Networks
Kevin Lee, Western Michigan University

CS49 - Computational Efficiency vs. Statistical Guarantee
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM
Grand Ballroom J

Organizer(s): Helen Zhang, University of Arizona
Chair(s): Helen Zhang, University of Arizona

10:05 AM
Embedding Learning
Xiaotong Shen, University of Minnesota
10:35 AM
Penalty Method for Variance Component Selection
Hua Zhou, UCLA
11:05 AM
Distributed Computing for Large Heteroskedastic Spatial Data
Zhengyuan Zhu, Iowa State University

CS50 - Developing Statistical Software For Drug Development
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM

Grand Ballroom K

Organizer(s): Yiming Peng, Genetech
Chair(s): Yiming Peng, Genetech

10:05 AM
Embrace R in Pharma - Building an R Community
Ning Leng, Genentech
10:35 AM
Reproducible Computation at Scale in R
Will Landau, Eli Lilly and Company
11:05 AM
Leveraging Open Source Tools for Drug Development
Douglas Kelkhoff, Genentech

CS51 - Machine Learning Problems in the Tech Industry
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM
Regency Ballroom AB

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

10:05 AM
Machine Learning Methods for Estimation and Inference in Differential Networks
Mladen Kolar, Chicago Booth
10:30 AM
Online and Offline Experimentation in Complex Systems
Akshay Krishnamurthy, .
10:55 AM
Modern recommendation systems: listwise collaborative ranking and non-stationary contextual bandits
James Sharpnack, UC Davis
11:20 AM
Discussant
Siva Balakrishnan, Carnegie Mellon University

CS52 - Grammar of Graphics: From Theory to Applications
Invited
Sat, Jun 1, 10:00 AM - 11:35 AM
Regency Ballroom C

Organizer(s): Jim Harner, West Virginia University

Chair(s): Zhi Yang, University of Southern California

10:05 AM
Unit Visualizations and the Grammar of Graphics
Steven Drucker, Microsoft
10:35 AM
ggplot2: An Extensible Platform for Publication-quality Graphics
Claus Wilke, University of Texas at Austin
11:05 AM
Tableau: Democratizing Visual Analytics by Automating Best Practices
Anushka Anand, Tableau

CS53 - The SAMSI Program on Model Uncertainty
Invited
Sat, Jun 1, 1:00 PM - 2:35 PM
Grand Ballroom I

Organizer(s): David Banks, Duke University / SAMSI
Chair(s): Dongchu Sun, University of Missouri

1:05 PM
The Stochastic Inverse Problem
Lei Yang, SAMSI
1:35 PM
Bayesian Model Calibration and Prediction Applied to Stochastic Simulators
Dave Higdon, Virginia Tech
2:05 PM
Uncertainty Quantification of Stochastic Computer Model for Binary Black Hole Formation
Derek Bingham, Simon Fraser University

CS54 - The IMS Program on Self-Consistency: a Fundamental Statistical Principle for Deriving Computational Algorithims
Invited
Sat, Jun 1, 1:00 PM - 2:35 PM
Grand Ballroom J

Organizer(s): Thomas Lee, UC Davis
Chair(s): Thomas Lee, UC Davis

1:05 PM
Likelihood-Free EM: Self-Consistency for Incomplete or Irregular-Pattern Data
Xiao-Li Meng, Harvard University

1:35 PM
Latent Variable Models, Self-Consistency, and Stochastic Approximation
Zhiqiang Tan, Rutgers University
2:05 PM
Self-Consistency as a Method to Develop Computationally Effective Algorithms for High-Dimensional Models
Alex Tsodikov, University of Michigan

CS55 - Recent Advances in Statistical Machine Learning and Reinforcement Learning
Invited
Sat, Jun 1, 1:00 PM - 2:35 PM
Regency Ballroom AB

Organizer(s): Will Wei Sun, University of Miami Business School
Chair(s): Hua Zhou, UCLA

1:05 PM
CORALS: Co-Clustering Analysis via Regularized Alternating Least Squares
Gen Li, Columbia University
1:35 PM
Model-Based Community Detection for Networks with Node Covariates
Ji Zhu, University of Michigan
2:05 PM
Nearly Optimal Adaptive Procedure with Change Detection for Piecewise-Stationary Bandit
Zheng Wen, Adobe Research

CS56 - Data for Human Health
Contributed
Sat, Jun 1, 1:00 PM - 2:35 PM
Grand Ballroom E

Chair(s): Xiyue Liao, Department of Statistics and Applied Probability, University of California, Santa Barbara

1:05 PM
Multiple-target Robust Design of a Coronary Stent with Multiple Functional Outputs
Fan JIANG, City University of Hong Kong
1:20 PM
Multiple Hypotheses Testing for Discrete Data - "MHTdicsrete" R package
Yalin Zhu, Merck & Co., Inc.
1:35 PM
What Are the Comorbidities That Go with Asthma? Basket Analysis Approach
Tianyuan Guan, University of Cincinnati

1:50 PM
An Optimal Kernel-Based U-Statistic Method for Quantitative Gene-Set Association Analysis
Tao He, San Francisco State University
2:05 PM
A Nonlinear Hierarchical Modeling Approach to Estimating the BAT Curve Using Markov Chain Monte Carlo
Colin O'Rourke, Benaroya Research Institute
2:20 PM
Floor Discussion

CS57 - Visualization Methods
Contributed
Sat, Jun 1, 1:00 PM - 2:35 PM
Regency Ballroom C

Chair(s): Tiffany Jiang, UC Davis

1:05 PM
Advanced Visualization Techniques for Big Data
Scott Lee Wise, SAS Institute, Inc.
1:20 PM
Interactive Ggplots in R
Zehao Xu, University of Waterloo
1:35 PM
Visualizing associations of multiple related but distinct phenomena
Maia P Smith, St George's University
1:50 PM
Data visualization techniques for the analysis of eczema-affected specific regions of the body as predictors of food allergy risk
Alyssa Ylescupidez, Benaroya Research Institute and the Immune Tolerance Network, Seattle
2:05 PM
Floor Discussion

Hackathon Update
Special Session
Sat, Jun 1, 1:00 PM - 2:35 PM
Regency Ballroom EF

Join the Hackathon participants as they present their findings.

CS58 - When Biomedical Data Gets Big: Challenges and Solutions in Biomedical Data Science
Invited

Sat, Jun 1, 2:45 PM - 3:50 PM
Grand Ballroom E

Organizer(s): James Eddy, Sage Bionetworks
Chair(s): Yalin Zhu, Merck & Co., Inc.

2:50 PM
Analysis of Whole Genome Sequence Analysis in >100k Individuals: Experience in the TOPMed Program
Ken Rice, Universiry of Washington
3:20 PM
Biomedical Informatics and Precision Medicine Are Laying the Framework for the Next Generation of Data-Driven Clinical Research
Sean Mooney, University of Washington

CS59 - Data Science Platforms: Docker and Kubernetes
Invited
Sat, Jun 1, 2:45 PM - 3:50 PM
Grand Ballroom I

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

2:50 PM
RsparkHub: Scaling Rspark with Kubernetes
Jim Harner, West Virginia University
3:20 PM
Using Rocker Containers and CI for Teaching R-Based Courses
Colin Wiiter Rundel, Duke University

CS60 - Expanding the Toolkit for Teaching Statistics
Invited
Sat, Jun 1, 2:45 PM - 3:50 PM
Regency Ballroom EF

Organizer(s): Alicia Johnson, Macalester College
Chair(s): Mikael Vejdemo-Johansson, CUNY College of Staten Island

2:50 PM
(A Picture-Book Approach To) Teaching the Analytics Process
Ruth M Hummel, SAS Institute / JMP Division
3:20 PM

Teaching Data Science Using Jupyter Notebooks and Binder
Brian Kim, University of Maryland

CS61 - Advances in Regression and Modeling
Contributed
Sat, Jun 1, 2:45 PM - 3:50 PM
Grand Ballroom J

Chair(s): Yongli Sang, University of Louisiana at Lafayette

2:50 PM
Nonparametric Estimation of a Mixing Distribution for Pharmacokinetic Stochastic Models
Alona Kryshchenko, California State University Cannel Islands
3:20 PM
Floor Discussion

CS62 - New Developments in Statistical Learning
Contributed
Sat, Jun 1, 2:45 PM - 3:50 PM
Regency Ballroom AB

Chair(s): Gen Li, Columbia University

2:50 PM
Flexible Functional Specification in Hierarchical Bayesian Estimation of Discrete Choices
Kali (Duke) Chowdhury, University of California, Irvine
3:05 PM
Correlation Tensor Decomposition and Its Application in Spatial Imaging Data
Yujia Deng, University of Illinois, Urbana-Champaign
3:20 PM
INDIVIDUALIZED MULTI-DIRECTIONAL VARIABLE SELECTION
Xiwei Tang, University of Virginia
3:35 PM
Quantile Regression for Big Data with Small Memory
Yichen Zhang, New York University

GS05 - Closing Keynote Address
General Session
Sat, Jun 1, 4:00 PM - 5:00 PM
Grand Ballroom E

Organizer(s): Kelly McConville, Reed College
Chair(s): Tim Hesterberg, Google

4:05 PM
Data Science and Statistics: Let's Not Call the Whole Thing Off!
Daniela Witten, University of Washington