

Printable Program
Printed on: 02/18/2019

Wednesday, May 29

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

Instructor(s): Garrett Golemund, 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 Golemund, 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

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

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

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

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

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.

Thursday, May 30

Welcome and Keynote Address
General Session
Thu, May 30, 9:15 AM - 10:30 AM

Organizer(s): Kelly McConville, Reed College

[Generalized Tensor Decompositions for Non-Normal Data](#)

Tamara Kolda, Sandia National Laboratories



Open Source and Community
Invited
Thu, May 30, 10:30 AM - 12:00 PM

Organizer(s): Gabriela de Queiroz, IBM

[Sustainers of the Tidyverse](#)

Mara Averick, RStudio

[Building a Community: The R-Ladies Story](#)

Gabriela de Queiroz, IBM

[Getting Involved in Scientific Open Source: Lessons from 7 Years of Growing the ROpenSci Community](#)

Karthik Ram, UC Berkeley



Deciphering Biological Systems via Innovative Statistical Learning Methods
Invited
Thu, May 30, 10:30 AM - 12:00 PM

Organizer(s): Tian Zheng, Columbia University

Chair(s): Tian Zheng, Columbia University

[Differential Network Connectivity Analysis](#)

Ali Shojaie, University of Washington

[Modeling Bias in Compositional Data](#)

David Clausen, University of Washington

[Extracting Biological Signals by Controlled Variable Selection](#)

Linxi Liu, Columbia University



Visual Storytelling

Invited

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

Organizer(s): Silas Bergen, Winona State University

Chair(s): Jerzy Wieczorek, Colby College

[What You Design Is Not What People See](#)

Alberto Cairo, University of Miami

[The Design and Evaluation of Expressive Visualization Tools for Data-Driven Storytelling](#)

Matthew Brehmer, Microsoft Research

[Things We've Learned from Telling the 'Fun' Data Stories](#)

Amber Thomas, The Pudding



Recent Developments in Lower Rank Learning for Complex Data

Invited

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

Organizer(s): Xiao-Li Meng, Harvard University

Chair(s): Raymond Wong, Texas A&M University

[Bayesian Analysis of the Covariance Matrix of a Multivariate Normal Distribution with a New Class of Priors](#)

Dongchu Sun, University of Missouri

[MCMC for Dempster-Shafer Statistical Inference](#)

Ruobin Gong, Rutgers University

[Deep Fiducial Inference](#)

Jan Hannig, The University of North Carolina at Chapel Hill



Teaching Statistics More Effectively to a New Generation of Students

Invited

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

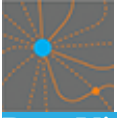
Organizer(s): Jo Hardin, Pomona College
Chair(s): Alejandra Castillo, Oregon State University

[Using GitHub with Statistics Undergraduates](#)

Jo Hardin, Pomona College

[Salt Fat Acid Heat: An Alternative to Cookbook Statistics](#)

Andrew Bray, Reed College



Data Visualization in Python

Invited

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

Organizer(s): Todd Iverson, Winona State University

Chair(s): Todd Iverson, Winona State University

[Introduction to Visualization with Python](#)

Stephen F. Elston, Quantia Analytics, LLC

[Altair: Declarative Visualization in Python](#)

Dominik Moritz, University of Washington



Project Jupyter

Invited

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

Organizer(s): Brian Granger, Cal Poly; Fernando Perez, UC Berkeley

[Sharing Reproducible Computations on Binder](#)

Lindsey J. Heagy, UC Berkeley

[Open Infrastructure in the Cloud with JupyterHub](#)

Chris Holdgraf, UC Berkeley

[JupyterLab: An Extensible and Flexible Platform for Collaborative Data Science](#)

Brian Ellison Granger, Cal Poly / Project Jupyter



Data Science's X-Factor

Invited

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

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

[Student Difficulties in Data Science Instruction: Early Findings](#)

Karl R. B. Schmitt, Valparaiso University

[Data Science In/Among/With/Toward the Humanities](#)

John Laudun, University of Louisiana

[Data Physicalizations: Where Art, Data, and Domain Applications Combine](#)

Katherine M. Kinnaird, Smith College



SADM Invited Papers

Invited

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

Organizer(s): Bertrand Clarke, University of Nebraska-Lincoln; Jia Li, Penn State University

[Bayesian Variable Selection in High-Dimensional EEG Data Using Spatial Structured Spike and Slab Prior](#)

Dipak K. Dey, University of Connecticut

[Mean Residual Function: a Tool for Exploring Patterns in Big Data](#)

Ehsan S. Soofi, University of Wisconsin-Milwaukee

[Model-Based Community Detection for Networks with Node Covariates](#)

Ji Zhu, University of Michigan



Communication Within and Beyond the Modern Data Science/Statistics Classroom

Invited

Thu, May 30, 4:00 PM - 5:30 PM

Organizer(s): Alicia Johnson, Macalester College

[Agile, Reproducible, and Accessible: Using Bookdown for Communication Within and Beyond the Classroom](#)

Alicia Johnson, Macalester College

[Using Slack for Communication and Collaboration in the Classroom](#)

Albert Y. Kim, Smith College

[Using Blogdown to Connect Beyond the Classroom](#)

Alison Hill, RStudio



Linguistic Diversity in NLP

Invited

Thu, May 30, 4:00 PM - 5:30 PM

Organizer(s): Rachael Tatman, Kaggle

[An Introduction to Computational Sociolinguistics](#)

Rachael Tatman, Kaggle

[English Isn't Generic for Language, Despite What NLP Papers Might Lead You to Believe](#)

Emily M. Bender, University of Washington

[Learning the Language of BlackTwitter](#)

Brandeis Hill Marshall, Spelman College



Shared Infrastructure for Data Science

Invited

Thu, May 30, 4:00 PM - 5:30 PM

Organizer(s): Soren Harner, Permaling

[The Machine Learning Lifecycle with MLflow](#)

Siddharth Murching, Databricks, Inc.

[Low-Latency Model Serving with MLflow and MLeap](#)

Corey Zumar, Databricks, Inc.



Recent Advances in Matrix and Tensor Factorization Models

Invited

Thu, May 30, 4:00 PM - 5:30 PM

Organizer(s): Raymond Wong, Texas A&M University

Chair(s): Jan Hannig, The University of North Carolina at Chapel Hill

[Low-Rank Covariance Function Estimation for Multidimensional Functional Data](#)

Raymond Wong, Texas A&M University

[Linked Matrix Factorization](#)

Eric F. Lock, University of Minnesota

[Boosted Sparse and Low-Rank Tensor Regression](#)

Kun Chen, University of Connecticut

Symposium on Data Science and Statistics Banquet

General Session

Thu, May 30, 6:30 PM - 8:00 PM

Organizer(s): Kelly McConville, Reed College

[Statistics Isn't All That Funny, but it Has Its Moments](#)

Joel Grus, Allen Institute for Artificial Intelligence
Friday, May 31

Friday Keynote Address
General Session
Fri, May 31, 8:30 AM - 10:00 AM

Organizer(s): Kelly McConville, Reed College

[Data Science: How the Union of Inferential Thinking and Computation Are Transforming Research and Education at Berkeley](#)

Fernando Perez, UC Berkeley



Advances in Analysis and Computing in Complex Data
Invited
Fri, May 31, 10:30 AM - 12:00 PM

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

[Graph-Based Change-Point Detection](#)

Hao Chen, UC Davis

[A Double Core Tensor Factorization and Its Applications to Heterogeneous Data](#)

George Michailidis, University of Florida

[Individualized Fusion Learning \(IFusion\) with Applications to Personalized Inference](#)

Minge Xie, Rutgers University



Data Science Platforms: Spark
Invited
Fri, May 31, 10:30 AM - 12:00 PM

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

[An R Interface to Hail](#)

Michael Lawrence, Genentech Research

[Scaling Sparklyr with Streams and Arrow](#)

Javier Luraschi, RStudio

[Using H2O in Spark with R](#)

Navdeep Gill, H2O.ai



A Field Guide to Education Tools in Data Science

Invited

Fri, May 31, 10:30 AM - 12:00 PM

Organizer(s): Alison Hill, RStudio

Chair(s): Alison Hill, RStudio

[Experiences in Teaching Data Science and Visualization at NASA](#)

David Meza, NASA

[Necessity Is the Mother of Invention: Evolution of a Data Science Team](#)

Adrienne Zell, Oregon Health and Science University



Recent Developments on Machine Learning

Invited

Fri, May 31, 10:30 AM - 12:00 PM

Organizer(s): Xiaotong Shen, University of Minnesota

Chair(s): Xiaotong Shen, University of Minnesota

[Shrinking Characteristics of Precision Matrix Estimators](#)

Adam J. Rothman, University of Minnesota

[P-Splines with an L1 Penalty for Repeated Measures](#)

Hui Jiang, University of Michigan

[Community Detection with Dependent Connectivity](#)

Annie Qu, University Illinois at Urbana-Champaign



Building and Growing Data Science Teams

Invited

Fri, May 31, 10:30 AM - 12:00 PM

Organizer(s): Jacqueline Nolis, Nolis, LLC

[From Zero to A^X: Scaling Data Science Teams](#)

Amanda Casari, Google Cloud

[Together at Last: Heterogeneous Teams and the Key to Success](#)

Heather Nolis, T-Mobile

[Creating Effective Data Science Teams](#)

Mehar Singh, ProCogia



Data Science Ethics Meet Reality
Invited
Fri, May 31, 1:30 PM - 3:00 PM

Organizer(s): Os Keyes, University of Washington

[The Politics of Data](#)

Os Keyes, University of Washington

[The Political Consequences of Repurposing Data](#)

Meg Young, University of Washington

[Beyond Methodological Rigor: Widening the Scope of Ethics in Data Science](#)

Anissa Tanweer, University of Washington



The Cutting Edge in Statistical Machine Learning
Invited
Fri, May 31, 1:30 PM - 3:00 PM

Organizer(s): Daniela Witten, University of Washington

Chair(s): Daniela Witten, University of Washington

[A Continuous-Time View of Early Stopping in Least Squares Regression](#)

Ryan Tibshirani, Carnegie Mellon University

[Fused Lasso on Graphs: Applications to Nonparametric Statistical Problems](#)

Oscar Hernan Madrid Padilla, UC Berkeley

[Two-Stage Computational Framework for Sparse Generalized Eigenvalue Problem](#)

Kean Ming Tan, University of Minnesota



Data Science Platforms: Deep Learning
Invited
Fri, May 31, 1:30 PM - 3:00 PM

Organizer(s): Javier Luraschi, RStudio

[Deep Learning and Probabilistic Programming with Applications to Intelligent Reality](#)

Soren Harner, Permaling

[R Interfaces to TensorFlow and Keras](#)

Kevin Kuo, RStudio

[Deep Learning Models at Scale with Apache Spark](#)

Joseph Kurata Bradley, Databricks, Inc.



Data Visualization Education

Invited

Fri, May 31, 1:30 PM - 3:00 PM

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

[Teaching Data Visualization: Integrating Theory and Practice](#)

Michael Freeman, University of Washington

[A Three-Part Data Visualization Curriculum](#)

Jerzy Wieczorek, Colby College

[Help Me Understand: Guiding Visualization Users with Annotations](#)

Robert Kosara, Tableau Software



Backend Data Science

Invited

Fri, May 31, 3:30 PM - 5:00 PM

Organizer(s): Edgar Ruiz, RStudio

[Data Science with Databases and R](#)

Edgar Ruiz, RStudio

[STOIC Next-Generation Spreadsheet: Bringing Data Science to the Masses](#)

Ismael Ghalimi, STOIC

[Working with Images and Text in R Through Embeddings](#)

Michael Lucy, Basilica



Computational Statistics for Large-Scale Biological Data

Invited

Fri, May 31, 3:30 PM - 5:00 PM

Organizer(s): Jacob Bien, University of Southern California

Chair(s): Kean Ming Tan, University of Minnesota

[Computationally Efficient High-Dimensional Interaction Modeling](#)

Guo Yu, University of Washington

[Inference for Diversity Under Networked Models](#)

Amy Willis, University of Washington

[Variance Component Testing and Selection for a Longitudinal Microbiome Study](#)

Jin Zhou, University of Arizona



Modern Multivariate Analysis

Invited

Fri, May 31, 3:30 PM - 5:00 PM

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

[The Multivariate Square Root Lasso: Computational and Theoretical Insights](#)

Aaron Molstad, Fred Hutchinson Cancer Research Center

[Estimating Multiple Precision Matrices Using Cluster Fusion Regularization](#)

Brad Price, West Virginia University

[\\$L_2\\$-Regularization and Some Path-Following Algorithms](#)

Yunzhang Zhu, The Ohio State University



Data Visualizations at the Institute for Health Metrics and Evaluation

Invited

Fri, May 31, 3:30 PM - 5:00 PM

Organizer(s): Brian Dart, IHME

[Building Interactive Data Visualization for a Global \(Health\) Audience](#)

Ryan Shackleton, University of Washington

[The Story of a Chart: Data Visualization Principles to Simplify Complexity](#)

Evan Laurie, University of Washington

[Behind the Scenes: Building Tools to Visualize Intermediate Results in Complex Data Science Pipelines](#)

Marlena Bannick, University of Washington



Democratizing Data Science with Workflows

Invited

Fri, May 31, 3:30 PM - 5:00 PM

Organizer(s): Michael I. Love, UNC-Chapel Hill

[Publishing Literate Programming Workflows in Scientific Journals](#)

Michael I. Love, UNC-Chapel Hill

[When Should You Add Github, Make and Docker to Your Data Science Workflow?](#)

Tiffany Timbers, University of British Columbia

[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



Interoperability: Your R Package Can Depend on Its Friends
Invited
Fri, May 31, 5:15 PM - 6:15 PM

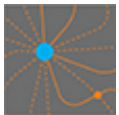
Organizer(s): Matthew N. McCall, University of Rochester

[Case Studies in Interoperability: From Generic Classes to Specific Functions](#)

Matthew N. McCall, University of Rochester

[How Core Data Structures Drive Interoperability in the Bioconductor Project](#)

Levi Waldron, CUNY SPH.



Grammar of Graphics: The Twentieth Anniversary
Invited
Fri, May 31, 5:15 PM - 6:15 PM

Organizer(s): Jim Harner, West Virginia University

[Past, Present, and Future of Grammar of Graphics Systems](#)

Lee Wilkinson, H2O.ai



Incorporating Ethics and Inclusion in Undergraduate Statistics Curriculum
Invited
Fri, May 31, 5:15 PM - 6:15 PM

Organizer(s): Brianna Heggeseth, Macalester College

[Ethics in an Advanced Undergraduate Seminar: Statistical Analysis of Social Network Data](#)

Miles Q. Ott, Smith College

[Intertwining Data Ethics into Intro Stats](#)

Brianna Heggeseth, Macalester College

Saturday, June 1



Machine Learning Problems in the Tech Industry

Invited

Sat, Jun 1, 10:00 AM - 11:30 AM

Organizer(s): Ryan Tibshirani, Carnegie Mellon University

Chair(s): Ryan Tibshirani, Carnegie Mellon University

[Machine Learning Problems in the Tech Industry](#)

Aaditya Ramdas, Carnegie Mellon University

[Machine Learning Methods for Estimation and Inference in Differential Networks](#)

Mladen Kolar, Chicago Booth

[Online and Offline Experimentation in Complex Systems](#)



Data Science for Fun

Invited

Sat, Jun 1, 10:00 AM - 11:30 AM

Organizer(s): David Smith, Microsoft

[Minecraft, R, and Containers](#)

David Smith, Microsoft

[Using Deep Learning in R to Generate Offensive License Plates](#)

Jacqueline Nolis, Nolis, LLC

[Fly Fishing with R](#)

Justin Shea, Roosevelt University



Computational Efficiency vs. Statistical Guarantee

Invited

Sat, Jun 1, 10:00 AM - 11:30 AM

Organizer(s): Helen Zhang, University of Arizona

Chair(s): Helen Zhang, University of Arizona

[Embedding Learning](#)

Xiaotong Shen, University of Minnesota

[Penalty Method for Variance Component Selection](#)

Hua Zhou, UCLA

[Distributed Computing for Large Heteroskedastic Spatial Data](#)

Zhengyuan Zhu, Iowa State University

680123689
790147900
462701581
462701581
358631009
680133689
.....

Data Science Platforms: Docker and Kubernetes

Invited

Sat, Jun 1, 10:00 AM - 11:30 AM

Organizer(s): Jim Harner, West Virginia University

[RsparkHub: Scaling Rspark with Kubernetes](#)

Jim Harner, West Virginia University

[Scalable R and Python Server Architectures Supporting Rich, Local Clients](#)

Alex Harner, Rc2.io

[Using Rocker Containers and CI for Teaching R-Based Courses](#)

Colin Wiiter Rundel, Duke University

680123689
790147900
462701581
462701581
358631009
680133689
.....

The SAMSI Program on Model Uncertainty

Invited

Sat, Jun 1, 1:00 PM - 2:30 PM

Organizer(s): David Banks, Duke University / SAMSI

Chair(s): David Banks, Duke University / SAMSI

[The Stochastic Inverse Problem](#)

Lei Yang, SAMSI

[Bayesian Model Calibration and Prediction Applied to Stochastic Simulators](#)

Dave Higdon, Virginia Tech

[Uncertainty Quantification of Stochastic Computer Model for Binary Black Hole Formation](#)

Derek Bingham, Simon Fraser University



Self-Consistency: a Fundamental Statistical Principle for Deriving Computational Algorithms

Invited

Sat, Jun 1, 1:00 PM - 2:30 PM

Organizer(s): Thomas Lee, UC Davis

Chair(s): Robin Gong, Rutgers University

[Likelihood-Free EM: Self-Consistency for Incomplete or Irregular-Pattern Data](#)

Xiao-Li Meng, Harvard University

[Latent Variable Models, Self-Consistency, and Stochastic Approximation](#)

Zhiqiang Tan, Rutgers University

[Self-Consistency as a Method to Develop Computationally Effective Algorithms for](#)

[High-Dimensional Models](#)

Alex Tsodikov, University of Michigan



When Biomedical Data Gets Big: Challenges and Solutions in Biomedical Data Science

Invited

Sat, Jun 1, 1:00 PM - 2:30 PM

Organizer(s): James Eddy, Sage Bionetworks

Chair(s): James Eddy, Sage Bionetworks

[Analysis of Whole Genome Sequence Analysis in >100k Individuals: Experience in the TOPMed Program](#)

Ken Rice, University of Washington

[Biomedical Informatics and Precision Medicine Are Laying the Framework for the Next Generation of Data-Driven Clinical Research](#)

Sean Mooney, University of Washington

[Global, Single Cell Atlases of Development](#)

Jay Shendure, University of Washington



Recent Advances in Statistical Machine Learning and Reinforcement Learning

Invited

Sat, Jun 1, 1:00 PM - 2:30 PM

Organizer(s): Will Wei Sun, University of Miami Business School

Chair(s): Will Wei Sun, University of Miami Business School

[CORALS: Co-Clustering Analysis via Regularized Alternating Least Squares](#)

Gen Li, Columbia University

[Slacked Empirical Likelihood for Computational Inference](#)

Yiyuan She, Florida State University

[Nearly Optimal Adaptive Procedure with Change Detection for Piecewise-Stationary Bandit](#)

Zheng Wen, Adobe Research

Closing Keynote Address

General Session

Sat, Jun 1, 4:00 PM - 5:00 PM

Organizer(s): Kelly McConville, Reed College

[Data Science and Statistics: Let's Not Call the Whole Thing Off!](#)

Daniela Witten, University of Washington

