

Invited Speaker

Peter DeWitt - University of Colorado School of Medicine

**Title:** Development of a Pediatric Blood Pressure Percentile Tool for Clinical Decision Support

**Abstract:** Pediatric blood pressure percentiles (PBPPs) are critical for diagnosis and several conditions, including hypertension and shock. PBPPs are a function of age, sex, and stature. Despite the known need for these values, many published tables for PBPPs are only applicable to children over one year of age and generally only provide values at and above the median. Further, the requirement of stature limits the utility further as stature is commonly missing from the patient records at pediatric ICU admission. The lack of readily accessible PBPPs is a barrier to patient care and the development of pediatric-specific clinical decision support tools. We developed and released the R package pedbp to fill this needed. Built on inputs from several PBPPs sources, pedbp provides PBPPs estimates based on age, sex and stature (known or unknown) and can be used to aid in the identification of children with hypertension and hypotension.

---

Evangeline Reynolds, West Point Military Academy

**Title:** A new wave of ggplot2 extenders

**Abstract:** There is a growing population of ggplot2 'super-users' that – given their regular and long-term ggplot2 use – have a strong working knowledge of grammar of graphics principles and of ggplot2's user interface. Among their ranks are statistics educators, data visualization professionals, and data scientists and statisticians.

I will argue that many ggplot2 super-users could find creating ggplot2 extensions very valuable, but that current educational materials are falling short in reaching and enabling many in this group. I'll then discuss how new material could be tailored to this group and will introduce some such new points of entry. I'll conclude by presenting how one new tutorial in this vein performed with a small group of statistics educators.

---

Joshua French, Associate Professor | Director of Data Science, University of Colorado Denver

**Title:** Prefiltered Component-based Greedy (PreCoG) Scan Method

**Abstract:** The spatial distribution of disease cases can provide valuable insights into disease spread and risk factors. Identifying disease clusters correctly can lead to the discovery of new risk factors and inform interventions that can help control and prevent the spread of disease. In this regard, we propose a novel scan method, the Prefiltered Component-based Greedy (PreCoG) scan method, which efficiently and accurately detects irregularly-shaped clusters using a prefiltered component-based algorithm. The PreCoG scan method is computationally efficient, flexible in its ability to detect irregularly-shaped clusters, while still being powerful and having high levels of sensitivity and positive predictive value. To demonstrate its efficacy, we compare its performance to many other scan-based methods. Additionally, we have included this method in the smerc R package to make it easy to apply this method to new data sets. Our proposed PreCoG Scan Method offers a unique and innovative approach to cluster detection that can improve the efficiency and accuracy of disease surveillance systems.

---

Jason Bernstein, Philip Myint (Lawrence Livermore National Laboratory), Melissa De Jesus (Florida International University)

**Title:** Memory-efficient and Online Generation of Equation-of-State Tables for Mixtures of Materials

**Abstract:** Simulations of industrial and natural phenomenon rely on properties of materials that are stored in equation-of-state (EOS) tables. These tables can each contain tens of thousands of values and are typically generated before the simulation is performed. However, when the simulation is of a mixture of multiple materials, it is not possible due to memory limitations to store enough tables to describe all possible mixture compositions. Hence, the need exists to represent the EOS behavior across composition space in a memory-efficient manner. In this talk, I discuss a statistical approach for generating EOS tables online, in which a small set of empirical basis functions are obtained from a set of existing tables and basis coefficients are predicted with Gaussian processes. Since predictions are needed for different mixture compositions, we use a Gaussian process kernel defined on the simplex, called the information diffusion kernel, to model correlation between tables. Prediction results are given for a two-component mixture consisting of iron and gallium. Time permitting, I will also discuss the task of calibrating EOS models (i.e., generating new EOS tables) from limited data. LLNL-ABS-852814 This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

---

Sven Serneels, Technology Director - Data Science, Aspen Technology

**Title:** Robustness-Inducing Transformations, with Application to Sparse PLS

**Abstract:** Robust alternatives exist for many statistical estimators. Many state-of-the-art robust methods are fine-tuned to optimize the balance between statistical efficiency and robustness. The resulting estimator may, however, require computationally intensive iterative procedures. Recently, several robustness-inducing transformations have been introduced. By merely applying such transformations as a pre-processing step, a computationally very fast robust estimator can be constructed. For the example of sparse partial least squares, we show that this can lead to performance close to the computationally more intensive methods.

---

Stephen Parziale, Yale University School of Medicine

**Title:** Implementing Biostatistical Methods within Sports Analytics using Propensity Score Matching

**Abstract:** The field of Biostatistics is a subset of Statistics that focuses on the study and analysis of biology. Biostatistical methods are designed to account for the “human error”, which attempts to adjust for the human element of data when performing analysis. Many of these techniques are applicable in the practice of Sports Analytics, another field that studies the behavior of human beings. Sports Analytics would be well suited to implement Biostatistical procedures that help in accounting for the subject matter being human beings. This project focuses on a new, but popular technique; Propensity Score Matching, and how it can be administered to improve the analysis of sports data

---

William Lippitt, Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus

**Title:** Quantifying disease through variograms: the textures of sarcoidosis

**Abstract:** Expert visual assessments of High Resolution Computed Tomography of the chest (chest CT) in patients with sarcoidosis have high inter- and intra-rater variation and are typically not clinically available. Reproducible quantitative radiomic measures associated with visual assessments of lung tissue texture and structure as well as other clinical outcomes are desirable in this context. Additionally, lungs, the organ of interest, vary widely in shape and size in healthy populations, with additional disease-relevant variation observed in patients with sarcoidosis. We investigate an alternative assessment of spatial structures in chest CT using location-specific variograms measuring spatial correlation. We then map these measures to visual assessments and other clinical outcomes and assess how these associations change according to changes in image processing pipeline, such as whether images are registered to a template prior to feature extraction and whether empirical variograms or variogram model fit parameters are used as measures.

---

Mallory Lai<sup>1</sup>, Macy Ricketts<sup>2</sup>, Naomi L. Ward<sup>2</sup>, and Tim Robinson<sup>1</sup>

<sup>1</sup>University of Wyoming, Department of Statistics, <sup>2</sup>Colorado State University, Department of Microbiology, Immunology, & Pathology

**Title:** Sources of Spurious Correlation in Microbiome Networks

**Abstract:** Correlation networks have risen in popularity as an intuitive way to illustrate potential ecological relationships between pairs of microbes<sup>1</sup>. Understanding the nature of pairwise ecological relationships among individual organisms is essential for defining microbial networks and requires carefully developed metrics of association. While metrics of correlation are standard fare in nearly every scientific field, their use for the discovery of potential microbial relationships has been called into question, with some authors asserting that correlation rarely provides reliable evidence for existing microbial interactions. Despite this, measures of correlation for defining microbial communities continues to be a widespread practice. More specifically, microbial networks built using correlation metrics can suffer from a phenomenon known as spurious correlation. Spurious correlation is a central issue, with many common sources, where the observed correlations are not reflective of any underlying biological process, and thus hold little to no predictive value. Here, we utilize simulations to illustrate the impact of spurious correlations on building microbiome networks. Spurious correlation is, in large part, due to the distinct characteristics of microbiome data. These data are high-dimensional, compositional, sparse, heterogeneous, and consist of causal relationships, which all possess their own unique statistical challenges. Awareness of the sources of spurious correlation, along with its fixes, is an important first step in guarding against improper use of correlation with microbiome data and improving reproducibility in microbiome research. The examples provided here highlight the critical need for future research to focus on limiting spurious correlation in microbiome data and tailoring our tools to handle the characteristics of these data.