Combining data visualization and statistical approaches for interpreting measurements and meta-data: Integrating heatmaps, variable clustering, and mixed regression models

Matthew A. Stiegel, Jon R. Sobus, Michael C. Madden, and Joachim D. Pleil

The advent of new higher throughput analytical instrumentation has put a strain on interpreting and explaining the results from complex studies. Contemporary human, environmental, and biomonitoring data sets are comprised of tens or hundreds of analytes, multiple repeat measures, stratification by fixed effects meta-data, and other parameters with unknown effects. We use an example of an environmental study wherein we measured airborne exposures, cytokines in blood, breath and urine, and standard medical parameters to illustrate how to sort, present, and interpret results and make them accessible for the ultimate audience, the general public. We feature the use of heatmaps and pattern recognition to help explain the complex underlying statistical relationships among variables and effects from gender, ethnicity and phenotype. We also demonstrate how visualized data can be used to optimize specific statistical evaluations and to inform future investigations.