As the amount of data being generated is exploding, we have entered the era of Big Data. To the extent that data can be analyzed, we may be able to gain a completely new perspective on our world, on how people interact, spend their resources, and organize their time. Though promises are held, the high-dimensionality and huge size of data sets can lead to inferential problems of their own—particularly spurious correlations, noise accumulation, and incidental homogeneity. New statistical thinking and computational approaches are required to handle these challenges. Many traditional methods that perform well for moderate sample size or low dimensional data do not scale to massive data or high dimensional data.

Our lab focuses on the development and application of advanced statistical models to analyze complex and high dimensional data (e.g. neuroimaging data). In particular, we have been focused on areas of feature selection, classification, regression and dimension reduction.