Reproducibility is imperative for any scientific discovery. More often than not, modern scientific findings rely on statistical analysis of high-dimensional data. At a minimum, reproducibility manifests itself in stability of statistical results relative to "reasonable" perturbations to data and to the model used. Jacknife, bootstrap, and cross-validation are based on perturbations to data, while robust statistics methods deal with perturbations to models. Moreover, the stability principle goes beyond data and model perturbations.

In this talk, a case is made for the importance of the stability principle in modern statistics and data science. Stability is not only necessary for interpretable statistical models, but also at the heart of CLT (a cornerstone of classical statistics) and modern limiting results. We put the principle into use to derive two new methods: ESCV (estimation stability with cross-validation) for regularization parameter selection in regression and staNMF (stability-driven non-negative matrix factorization) for number of component selection. ESCV and staNMF's successes are then demonstrated with movie reconstruction and Drosophila embryonic image data, respectively.