The decade-long Sloan Digital Sky Survey produced detailed spectral observations of over a million distant galaxies: a dataset that is still yielding new insights years after its release. The data are very high-dimensional: observations of each galaxy include fluxes measured at nearly 4000 distinct wavelengths. Astronomers have long employed dimensionality reduction algorithms such as PCA to understand the structure of this dataset, but the nonlinear nature of the structure means such linear transformations miss important features, and nonlinear manifold learning approaches, though promising, have historically been too slow to be applicable on the entire dataset. Here I will report on some initial exploratory work with the new megaman package for scalable manifold learning (introduced in another talk by my colleague Marina Meila), particularly the approaches we have used to apply such algorithms to noisy and incomplete observations.