Practice some of the initial steps of data analytics
Data sources of mini project #1 or find new, data should be large!
Use R or python for processing, D3 for visualization
  - note: shiny only works on windows right now
  - deliver a report and URL that allows selection of all options
Data decimation (20 points)
  - compare random sampling with adaptive sampling
Dimension reduction (use decimated data) (10 points)
  - find the intrinsic dimensionality of the data using PCA
Visualization (use dimension reduced data) (40 points)
  - compare PCA, MDS (Euclidian, cosine, correlation), and isomap

Extra credit: text visualization (30 points)
  - stem, remove stop words, perform tf-idf, then LSA, and visualize