We strongly suggest to work in groups of up to 3. Due date Tu 1-30-18.

1. Pick a dataset of your choice (with at least 1000 data points) and apply the following dimensionality reduction techniques:
   - t-SNE (paper, code)
   - LargeVis (paper, code)
   - TriMap (paper, code)
   - Optional: PCA, MDS, LLE, Isomap, etc. (MATLAB code)

You can use the datasets from UCI repository, Kaggle, etc. or data related to your own research. Provide a scatter plot of each embedding. You can color each data point based on class labels, certain features, etc.

2. Implement the following quality measures of dimensionality reduction and compare the results of the different methods numerically (you only need to implement 2 methods). Three examples are given below:
   - Mean precision-recall [1]
   - Trustworthiness-continuity [2]
   - Optional: Nearest neighbor accuracy (for labeled datasets)

Are the results conclusive? Can you think of better ways to assess the quality of the embedding (for instance, using domain knowledge)?

3. Finally, challenge yourself to find something additional: A new invariance that some of the DR methods satisfy, improve t-SNE by flipping the relative entropy, ...

Hints:
   - We recommend implementing all the code in Python.
   - First find some group members and partition the tasks.
References
