Lecture 18 Outline:

1. Review Matrix Factorizations / Dimension Reduction.

2. Clustering (data segmentation; finding groups).
   - Minimize the within-cluster dissimilarity.
   - Dissimilarity measures.
   - Hard vs. Soft clustering.
   - Clustering variables vs. cluster observations vs. clustering both (bi-clustering).

   - Minimizes the within cluster dissimilarity based on Euclidean distance.
   - $K$-means algorithm.
   - Properties.
   - Simulated examples.

4. Choosing the number of clusters for $K$-means:
   - Heuristic methods - look for the “kink” in the loss function.
   - Gap statistic.
   - Silhouette Statistic.
   - Prediction Strength.
   - Cluster Stability.

5. Intro to Spectral Clustering.