Clustering

- **Clustering**: Intuitively, finding clusters of points in the given data such that similar points lie in the same cluster

- Can be formalized using distance metrics in several ways

  - Group points into k sets (for a given k) such that the average distance of points from the centroid of their assigned group is minimized

    - Centroid: point defined by taking average of coordinates in each dimension.

  - Another metric: minimize average distance between every pair of points in a cluster

- **Hierarchical clustering**: example from biological classification

  - (the word classification here does not mean a prediction mechanism)