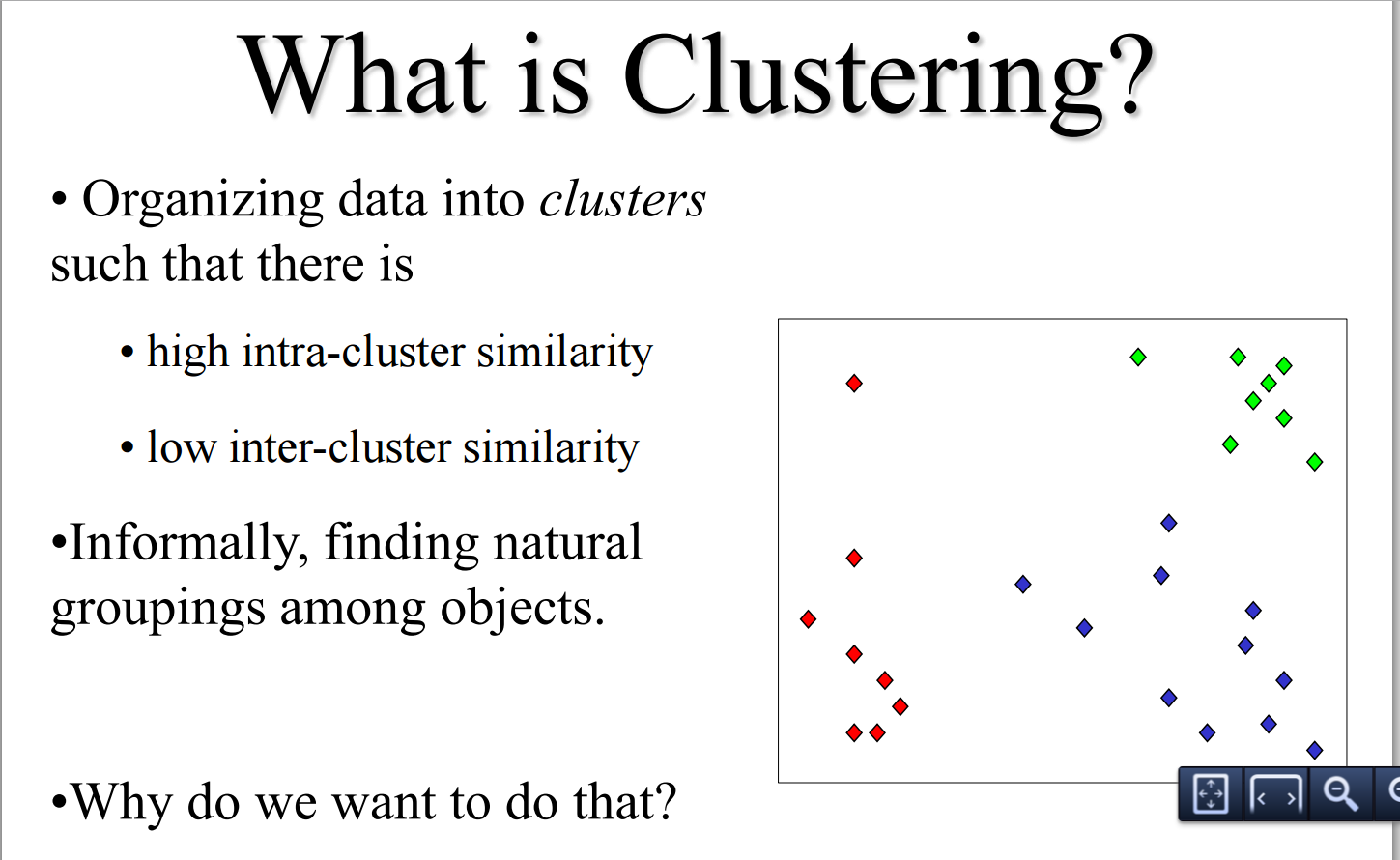
**Unsupervised learning** is a type of machine**learning** algorithm used to draw inferences from datasets consisting of input data without labeled responses. The most common **unsupervised learning** method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data.

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), a **feature** is an individual measurable property of a phenomenon being observed.

The initial set of raw features can be redundant and too large to be managed. Therefore, a preliminary step in many applications of [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition) consists of [selecting](https://en.wikipedia.org/wiki/Feature_selection) a subset of features, or [constructing](https://en.wikipedia.org/wiki/Feature_extraction) a new and reduced set of features to facilitate learning, and to improve generalization and interpretability. **Cluster analysis** or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters).



Hierarchical Clustering:

Bottom-Up (agglomerative): Starting with each item in its own cluster, find the best pair to merge into a new cluster. Repeat until all clusters are fused together. We begin with a distance matrix which contains the distances between every pair of objects in our database.

**Principal component analysis** (**PCA**) is a statistical procedure that uses an [orthogonal transformation](https://en.wikipedia.org/wiki/Orthogonal_transformation) to convert a set of observations of possibly correlated variables into a set of values of [linearly uncorrelated](https://en.wikipedia.org/wiki/Correlation_and_dependence) variables called **principal components**. The principal components are orthogonal because they are the [eigenvectors](https://en.wikipedia.org/wiki/Eigenvector) of the [covariance matrix](https://en.wikipedia.org/wiki/Covariance_matrix), which is[symmetric](https://en.wikipedia.org/wiki/Symmetric_matrix#Real_symmetric_matrices). PCA is sensitive to the relative scaling of the original variables.

K means:

Given a set of observations (**x**1, **x**2, …, **x***n*), where each observation is a *d*-dimensional real vector, *k*-means clustering aims to partition the *n* observations into *k* (≤ *n*) sets **S** = {*S*1, *S*2, …, *Sk*} so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find:

\underset{\mathbf{S}} {\operatorname{arg\,min}}  \sum_{i=1}^{k} \sum_{\mathbf x \in S_i} \left\| \mathbf x - \boldsymbol\mu_i \right\|^2 

where ***μ****i* is the mean of points in *Si*.