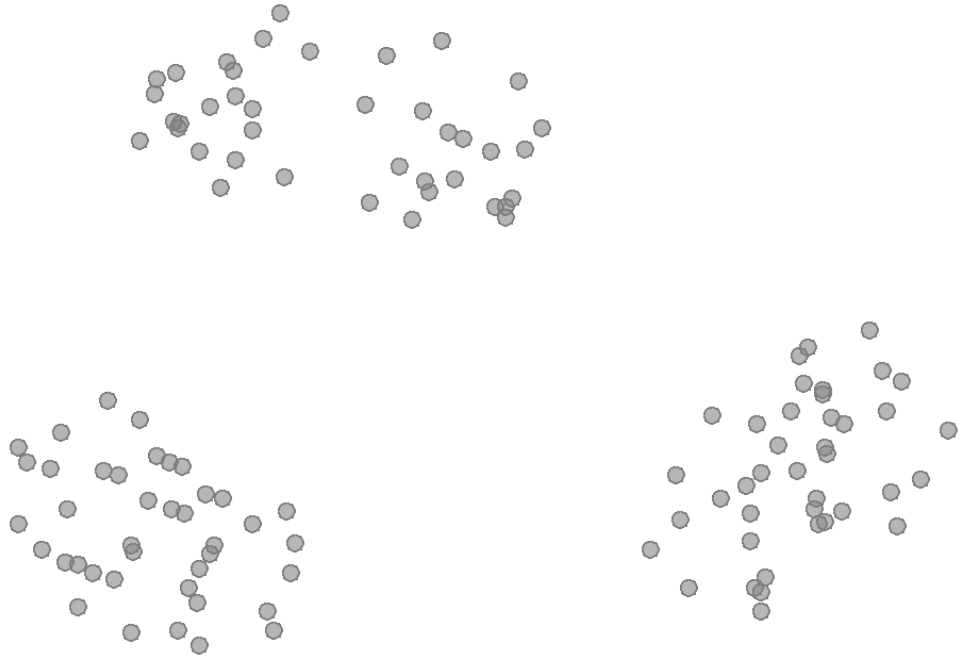
## Clustering (Unsupervised)

The goal of clustering is to automatically group similar samples into sets.

Since the clustering algorithm does not have any advanced knowledge of how the sets are to be defined, and furthermore, since this entire process is *automatic*, the clustering algorithm must have a metric of gauging sample-similarity. It does this the same way we, as humans, do: by looking at the various characteristics and features of the sample.



**More Examples**

- Match similar individuals on a matrimonial site based on their answer to profile questions.
- Based on their search history, recommend houses a prospective home-buyer might be interested in considering.
- Pinpoint the most likely location for a future earthquake using past earthquake seismic data.
- Identify new characteristics shared by different people suffering from the same disease.

There are different types of clustering algorithms, some supervised, some unsupervised. There are even semi-supervised clustering methods as well. In this course, you'll be dealing only with unsupervised clustering because you won't be providing the computer with anything except your raw

data. No labels will be shared that tell the computer what the *correct* cluster assignment is.