Franzen (2008) provides the examples of the deterministic clustering. One widely used *deterministic* method involves *hierarchical clustering*. It starts with as many clusters as there are observations, and the number of clusters is decreased one by one, at each step. Two groups are merged at each stage, according to certain optimization criteria. Commonly used criteria for merging are cluster measures such as smallest dissimilarity (single-linkage), average dissimilarity (average linkage), or maximum dissimilarity (complete linkage). In single linkage, the distance between two clusters is represented by the minimum distance between all possible pairs of objects. In average linkage, the distance used is the average of all pairs of objects and complete linkage is based on the maximum distance between all possible pairs of objects in the two clusters. Wardís method is another *hierarchical* method. It forms clusters by maximizing within-cluster homogeneity. The measure of homogeneity is the within-group sum 1 of squares. The method tries to minimize the total sum of squares by in each step merging the two clusters for which the increase of the sum of squares are the lowest. Wardís method creates clusters of near equal size, having close to hyper spherical shapes (Franzen 2008). Another commonly used *deterministic* method is *non-hierarchical clustering*, which is based on iterative relocation. These methods do not create a tree structure to describe the groupings in data, but create rather a single level of clusters. Objects are relocated between a predetermined number of groups until there is no further improvement according to the criteria used. As opposed to hierarchical clustering, the number of groups must be known prior to the clustering. *K-means* clustering is a non-hierarchical clustering algorithm (Franzen 2008). As an example of non-hierarchical deterministic clustering may serve K-means (unsupervised generative model). *K-means* clustering (Nakashe 2018) is a non-hierarchical clustering algorithm which uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible (Franzen 2008). k-means clustering is unsupervised algorithm because data points are unlabeled and it learns the cluster distribution, not the cluster boundaries. It may also be treated as generative as it learns distributions of clusters, not boundaries of the clusters.

*Deterministic clustering* is suited for cohesive and well-separated groups, but is not constructed for clusters with different geometric forms, nor for situations with overlapping groups. Moreover, these methods are not based on standard principles of statistical inference and do not provide an assessment of clustering uncertainties.

Landau (2010) states that most cluster analysis methods are essentially heuristic methods in the sense that they do not make explicit assumptions about the data-generating process. It is therefore impossible to infer from sample to population. Perhaps this presents no real difficulties to investigators involved in an initial exploration of their data where cluster analysis is only used to suggest hypothesis for future investigation. However, attempts have been made to develop a more acceptable statistical approach to the clustering problem, using what are known as finite mixture distributions (Landau 2010).

Following Franzen (2008), *model-based cluster analysis* is another cast of mind developed in recent years which provides a principled statistical approach to clustering. The idea is to base cluster analysis on a probability model. The population of interest consists of different subpopulations, each with its own distribution. Data is viewed as coming from a mixture model where each distribution represents a cluster. The overall population is a mixture of these subpopulations. The resulting model is a finite mixture model. The development of cluster analysis in this direction opens for understanding of the true process and origin of clusters (Franzen 2008). Various geometric properties are obtained through different parametrization of the distributions, or even completely different distributions among clusters. Measurement errors are an inherent part of the model, and outliers can be modeled by adding a distribution with larger variance or a different distribution than the rest of the clusters in the mixture (Franzen 2008).