3.2 Clustering

We now proceed with clustering methods. As previously discussed, clustering refers to the division of data into groups. As before, we are given $N \in \mathbb{N}$ data points $\mathcal{D} := \{x_i\}_{i=1}^N \subset \mathbb{R}^d$. We also assume a fixed number $K \in \mathbb{N}$ of groups (clusters?)

We begin with a wish list. We are looking for:

1'. **Assignment of data points to cluster centers**: $x_i \mapsto k_i \in \{1, \ldots, K\}$ for all $i = 1, \ldots, N$

1. General assignment rule: $x \mapsto k(x) \in \{1, \ldots, K\}$ for all $x \in \mathbb{R}^d$
2. Reconstruction rule: $k \mapsto m_k \in \mathbb{R}^d$ for all $k \in \{1, \ldots, K\}$

We note that 1 is a special case of 1. Also notable is 2, where we seek a kind of inverse mapping (more precisely, a right inverse), which assigns a representative element m_j to each group k. We also refer to m_j as the **mean** of the k-th group. It is also reasonable to choose this mapping so that m_k belongs to the k-th group.

Next up: