

How Many Clusters should we Use with k -means ?

In practice

- no supervision from data \Rightarrow no way to automatically choose k
- heuristics and "clustering quality measures" all resort on assumptions
- there is **no** "natural number of clusters" (expect for toy problems)
- the choice of k depends on what the user wants to do with data

Interaction with the user

- clustering is mainly used to "explore" data
- interaction with the user is **necessary**
- in some cases, clustering is only a preprocessing step \Rightarrow supervision ?