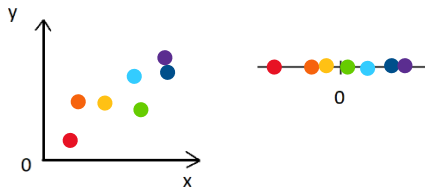


## Třídění dat do skupin

Neboli **shluková analýza** třídí objekty do skupin buď hierarchicky, přičemž vzniká tzv. **dendrogram**, nebo nehierarchicky, např. **metodou k-průměrů**, jejímž cílem je nalezení takových skupin, kdy vnitroskupinová podobnost je co největší a zároveň jsou skupiny od sebe co nejvíce odděleny. Vzhledem k tomu, že počet skupin je nutno definovat předem, je užitečné vyzkoušet několik počtů skupin a analýzou rozptýlů najít největší poměr „variabilita mezi skupinami / variabilita uvnitř skupiny“.

Čím víc proměnných (dat) definuje objekt (vzorek), tím spolehlivější jsou výsledky, ale tím komplikovanější je zpracování. Někdy je objekt definovaný tolika proměnnými, že je dobré před shlukováním jejich počet omezit. Původní proměnné je možné transformovat na menší počet (při zachování většiny informace) procesem **extrakce proměnných**. Mezi extrakční metody patří **Ordinační analýzy**, např.:

**Analýza hlavních komponent (PCA)** seskupí proměnné, které spolu lineárně korelují, do tzv. hlavních komponent. Komponenty jsou pak seřazeny podle toho, jak moc se podílí na zachování variability dat (kolik informace užitečné pro shlukování nesou).



Pozn.: pokud máme předem definované skupiny a chceme do nich vzorky rozřídít, použijeme některou z **klasifikačních metod** (Bayesův klasifikátor, Metoda k-nejbližších průměrů a Metoda podpůrných vektorů).