



PSY532: Dendextend

Tomáš Kratochvíl



Vymezení

- CO TO JE?
 - rozšíření pro clusterovou analýzu
 - porovnání 2+ dendrogramů
 - nalezení „nejhezčího“ dendrogramu
 - úpravy vzhledu dendrogramu jako:
 - (1) otáčení dendrogramu a clusterů
 - (2) tvorba labelů a jejich obarvení
 - (3) zabarvení shluků do určité vzdálenosti
+ úpravy těchto barev
 - tj. umožňuje **lepší vizualizaci a výběr typu** clusterové analýzy
... a další práci s daty jako korelační analýzy clusterů,
porovnání clusterových analýz, přídatné ploty k dendrogramům aj.

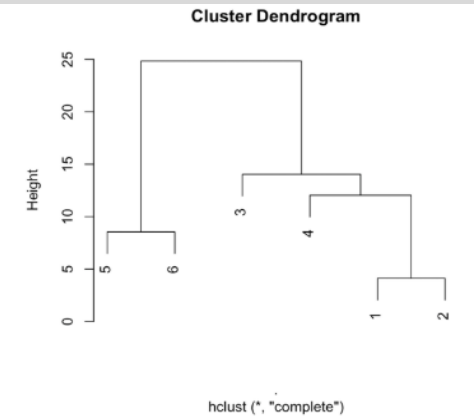
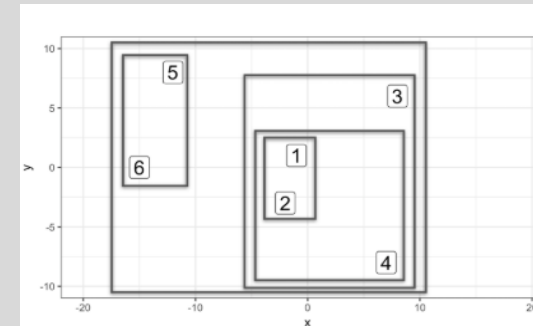
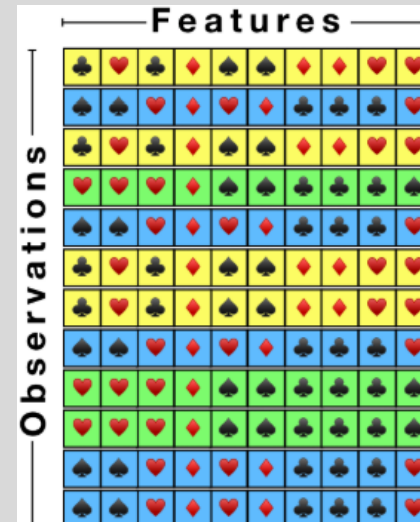
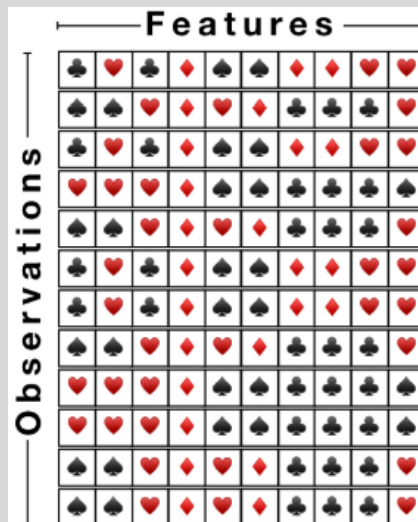
Jak to vlastně vypadá?

- CO JE „SHLUKOVÁNÍ DAT“?

- explorativní analýza
- podobná pozorování, výrazně odlišná od ostatních
- řeší se na základě vzdálenosti

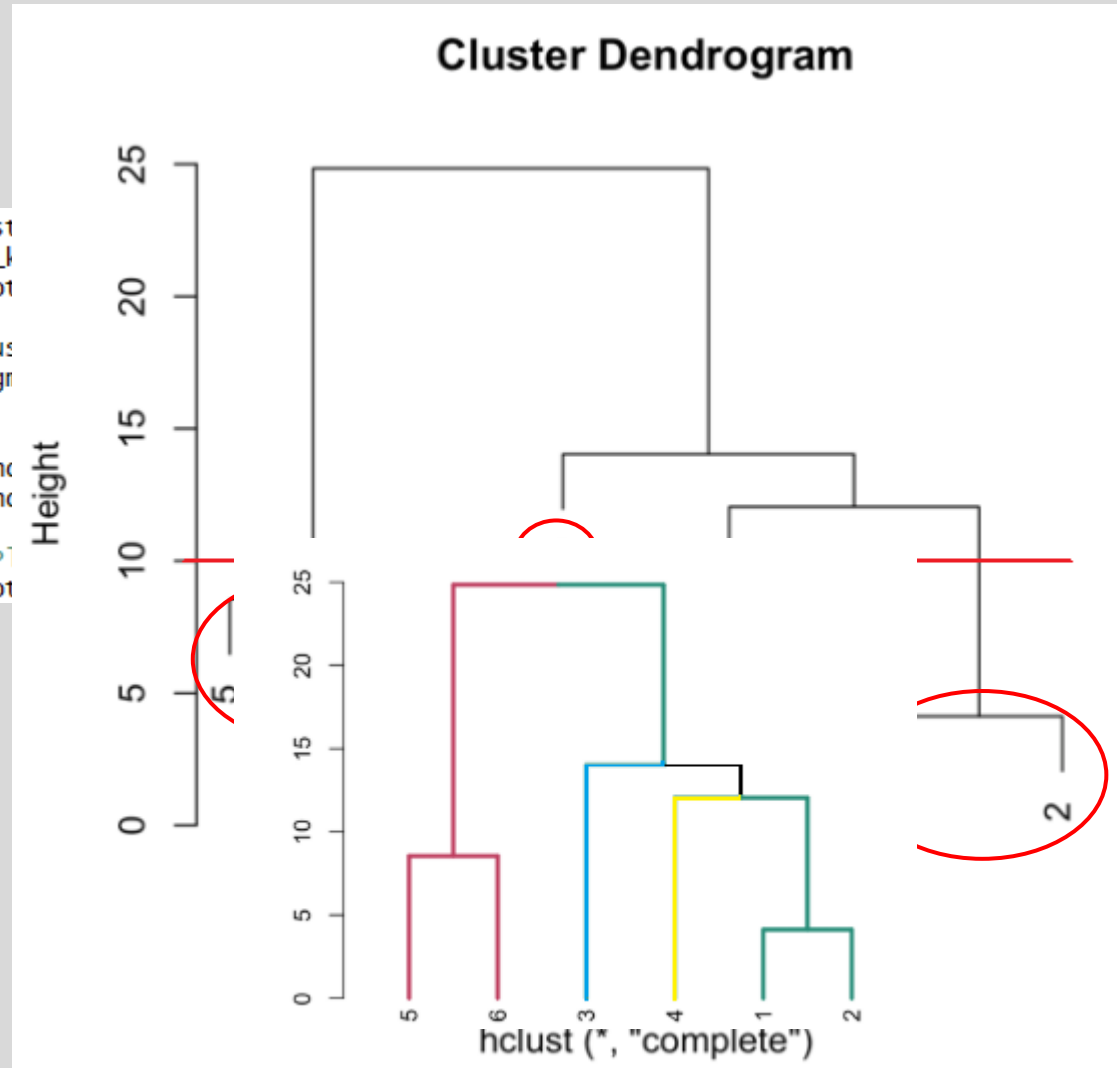
- CO JE „DENDROGRAM“?

- vizualizace iterativního procesu shlukování dat
- zachycuje vzdálenost, v jaké vzniká další shluk



Příkazy a jejich funkce

```
58 dist
59 hc_l
60 plot
61
62 clus
63 segr
64
65
66 dend
67 dend
68
69 # P
70 plot
```



- lean") → vytvoření vzdáleností
-) → vytvoření shluků na základě max. vzdáleností
-) → oddělení shluků na základě vzdálenosti „10“
- převedení na dendrogram
- obarvení větví, které spadají do jednoho shluku po vzdáleností „10“

Příkazy a jejich funkce

Vymezení

Co a jak

Příkazy
&
Funkce

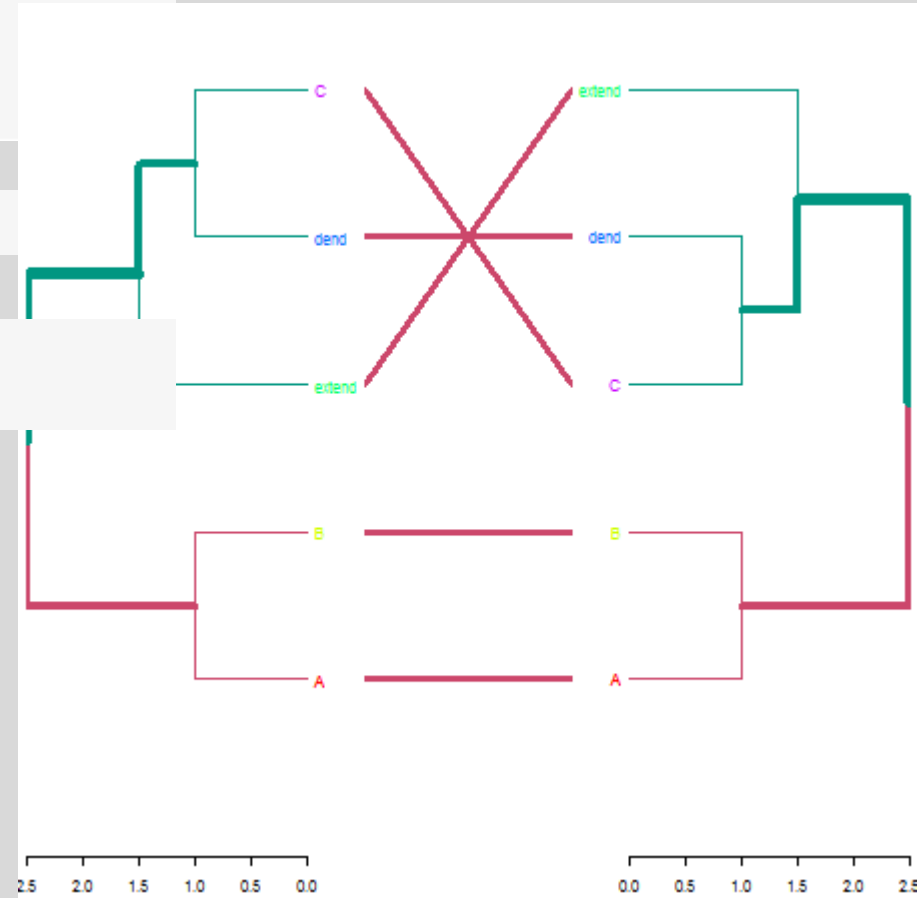
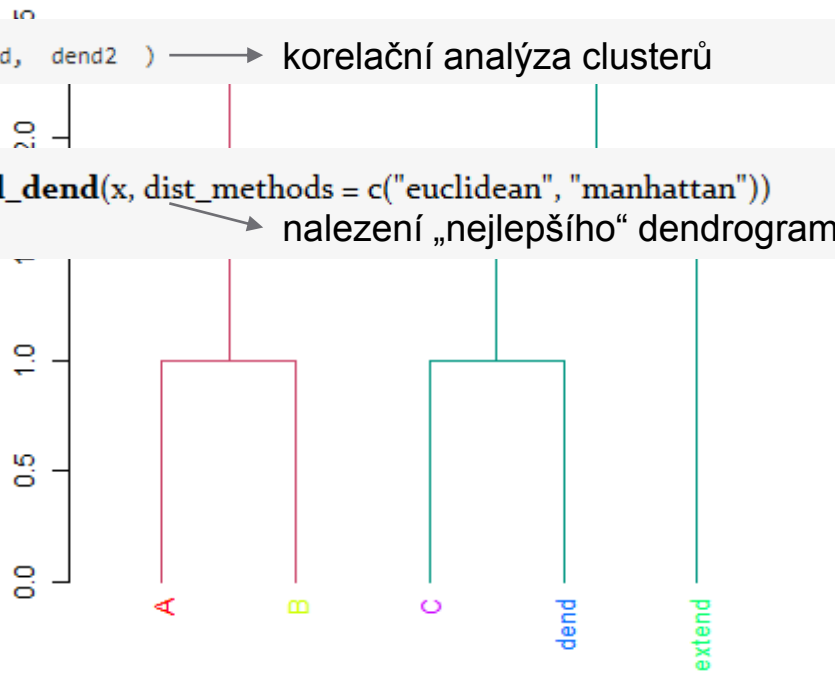
```
labels(dend)           → přidá labely ke clusterům
labels_colors(dend) <- rainbow(5)
labels_colors(dend)    → obarví labely

dend2 <- sort(dend)    → seřadí clustery dle vzdálenosti a velikosti
plot(dend2)

tanglegram(dend, dend2) → porovná dendrogramy
```

```
cor_cophenetic(dend, dend2) → korelační analýza clusterů
```

```
best_dend <- find_dend(x, dist_methods = c("euclidean", "manhattan"))
plot(best_dend) → nalezení „nejlepšího“ dendrogramu
```



PSY532: Dendextend

Tomáš Kratochvíl

Díky za pozornost!

