



**GIS4SG**

**Prostorové analýzy II.**

**podzim 2023**

**Lukáš Herman**

[herman.lu@mail.muni.cz](mailto:herman.lu@mail.muni.cz)

**Laboratory on Geoinformatics and Cartography (LGC)  
Institute of Geography  
Masaryk University  
Czech Republic**

# Dnešní úkoly

- **Vyzkoušejte si prostorovou analýzu plošných dat** v prostředí programu GeoDa
- Vstupní data (v ISu):
  - Obecní byty v Brně



# SHLUKOVÁNÍ – APLIKACE

- **Redukce množství dat a průzkumu multidimenzionálního atributového prostoru** s cílem identifikovat malý počet zajímavých subdimenzí (resp. kombinací atributů), které pak mohou být zkoumány z prostorového hlediska (uplatnění klasických multivariačních metod a následně vizualizace výsledků a jejich interpretace).
- **Průzkumu prostorových vzorů) a vztahů.**
- **Prostorová klasifikace a diskriminace („rozdělování“).**



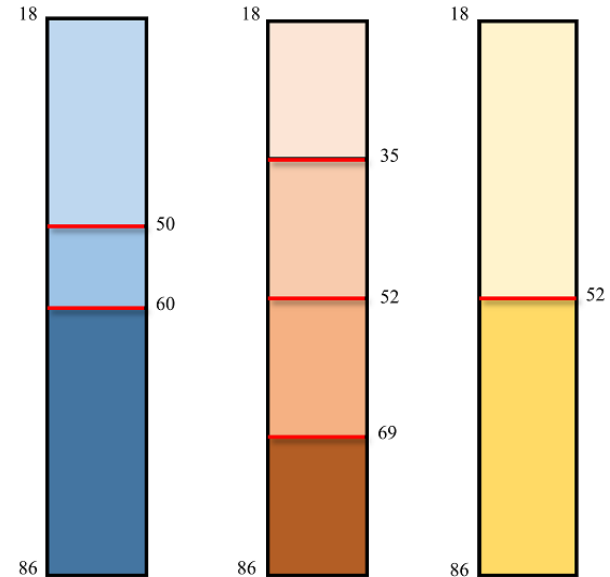
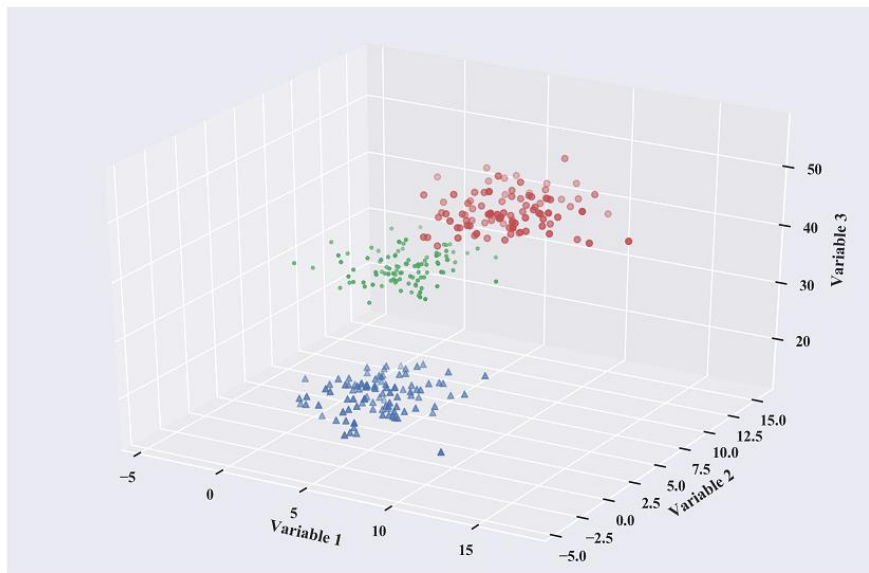
# SHLUKOVÁNÍ

- **Shluková analýza** je společný název pro celou řadu metod, jejichž cílem je využití informací z analýzy vícerozměrných dat k roztrídění množiny objektů do několika relativně homogenních podsouborů, označených jako shluky (**clustery**).
- Objekty uvnitř shluků mají být co nejvíce podobné a objekty patřících do různých shluků co nejvíce rozdílné. Podobnost mezi objekty je uplatněna jako kritérium pro tvorbu shluků objektů.
- Podobnost se měří různými prostředky:
  - míry korelace – korelační koeficienty (Pearsonův, Spearmanův)
  - míry vzdálenosti – euklidovská vzdálenost, Manhanattanská vz., ...
  - míry asociace – *nominální (kvalitativní) data* – Sokalův-Michenerův koeficient asociace, Russelův-Raoův koeficient asociace, ...
  - Korelační a vzdálenostní míry jsou míry metrických dat



# ROZDĚLENÍ METOD – počet proměnných

– UNIVARIATE versus MULTIVARIATE



(A) Manual Classification with 2 class breaks

(B) Equal Interval Classification with 4 class breaks

(C) Defined Interval Classification with 2 class breaks

Assign point to the closest cluster (A)



Assign all points to the closest cluster (B)





# ROZDĚLENÍ METOD II. – princip

Tab. 17-1 Základní rozdělení metod shlukové analýzy

Skupina	Metoda	Poznámka
Hierarchické	aglomerační (sdružovací)	Postupným seskupováním vytváří stromovou strukturu od jednotlivých objektů až po 1 shluk
	divizní (rozdělovací)	rozdělují počáteční celkový shluk do hierarchického systému dílčích skupin či objektů
Nehierarchické	optimalizační	
	analýzy modů	



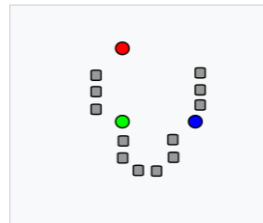
# ROZDĚLENÍ METOD IIO. – „prostorovost“

- Lze pracovat jen s atributy! (bez prostorové složky) → Statistika apod.
- „Soft“ prostorové
  - K-means
  - DBSCAN
- „Hard“ prostorové
  - SKATER = Spatial `K`luster Analysis by Tree Edge Removal,
  - REDCAP = REgionalization with Dynamically Constrained Agglomerative clustering and Partitioning

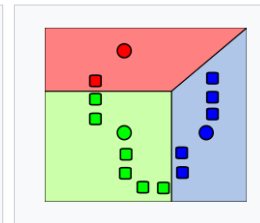
# K-means

- Metoda „k průměrů“
- algoritmus **nehierarchické** shlukové analýzy.
- Předpokládá, že shlukované objekty lze chápat jako body v nějakém eukleidovském prostoru a že počet shluků  $k$  je předem dán (případně lze vyzkoušet různá  $k$ , pro každé spustit algoritmus znovu a výsledky porovnat).
- Shluky jsou definovány svými centroidy
- Objekty se zařazují do toho shluku, jehož centroidu jsou nejbližší.
- Algoritmus postupuje iterativně tak, že se vyjde z nějakých (obvykle náhodně zvolených) centroidů, přiřadí do nich body, přepočítá centroidy tak, aby šlo o těžiště shluku bodů, pak opět přiřadí body k nově stanoveným centroidům a tak dále, až dokud se poloha centroidů neustálí.

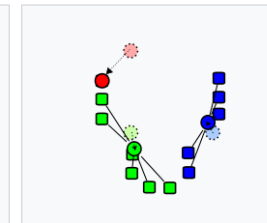
Demonstrace algoritmu



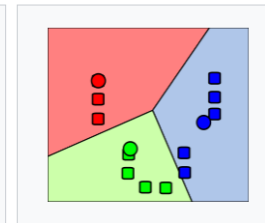
1.  $k$  výchozích centroidů (zde je  $k=3$ ) se náhodně umístí v prostoru dat (shlukované objekty šedé, centroidy barevné)



2. Objekty se přiřadí nejbližším centroidům, čímž vznikne  $k$  shluků. Centroidy tak definují **Voroného teselaci** prostoru.



3. Přepočtenou se centroidy shluků tak, aby šlo o těžiště objektů, jež patří do těchto shluků.



4. Kroky 2 a 3 se opakují, dokud nedojde k ustálení (**konvergence**).



# K-means

KMeans Clustering Settings

Input:

Select Variables

- COUNT
- AVE\_ID\_
- dept
- Crm\_prs
- Crm\_prp
- Litercy
- Donatns
- Infants
- Suicids
- MainCty
- Wealth
- Commerc

Use geometric centroids: Auto Weighting  
Weighting: 0 — 1 1

Select Spatial Weights:

Parameters:

Number of Clusters: 5

Minimum Bound:  const 1  
10%

Transformation: Standardize (Z)

Initialization Method: KMeans++

Initialization Re-runs: 150

Use Specified Seed:  Change Seed

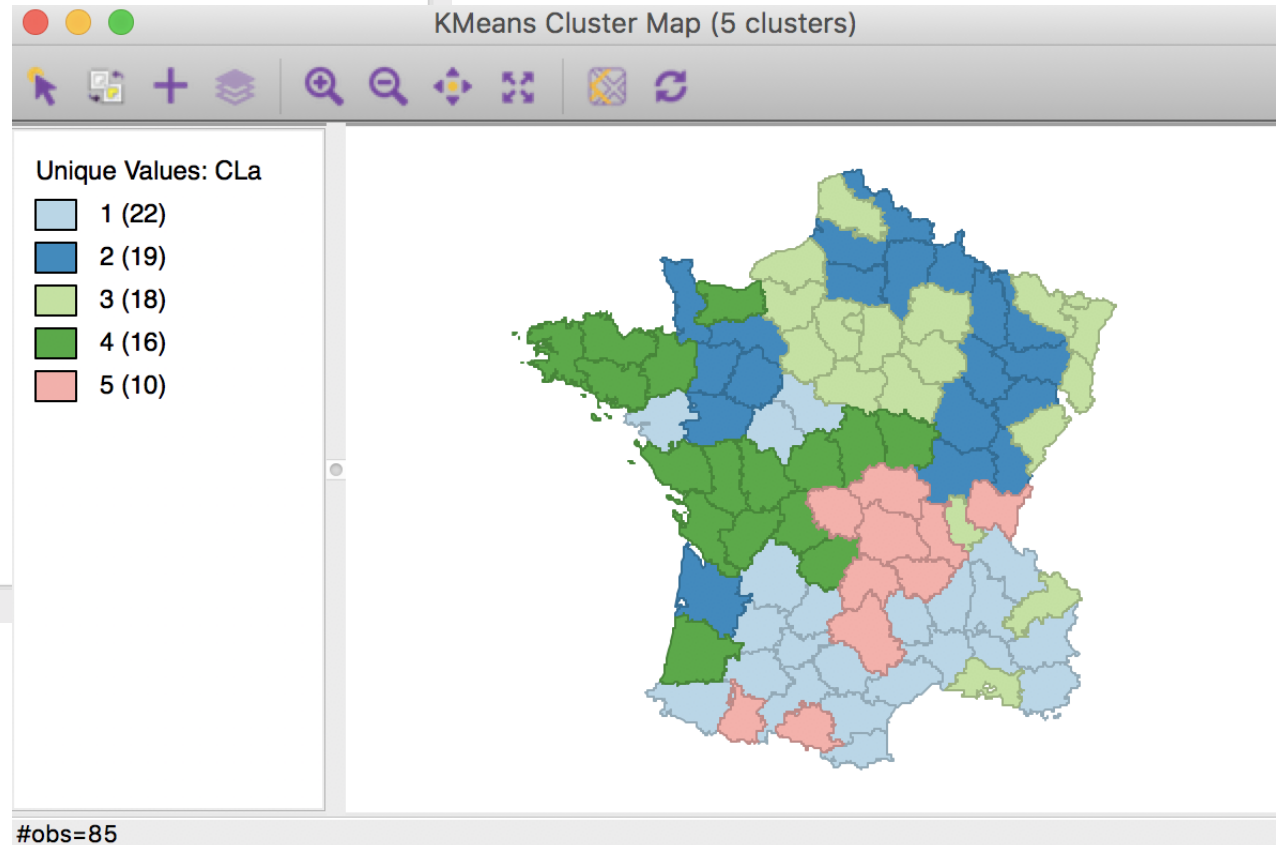
Maximum Iterations: 1000

Distance Function: Euclidean

Output:

Save Cluster in Field: CLa

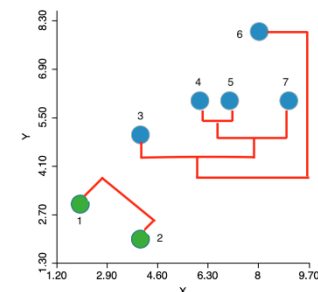
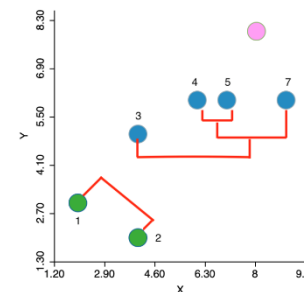
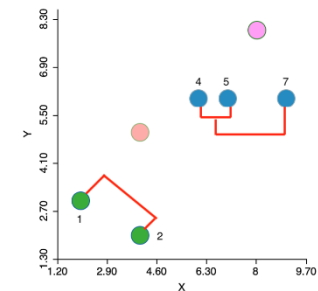
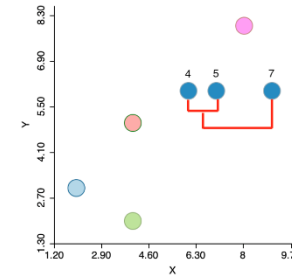
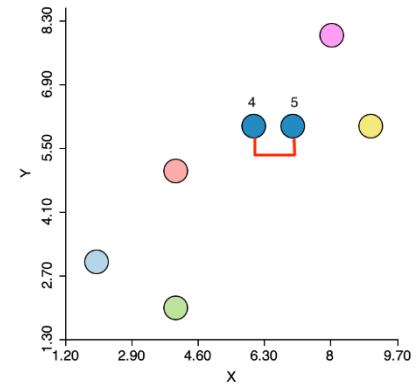
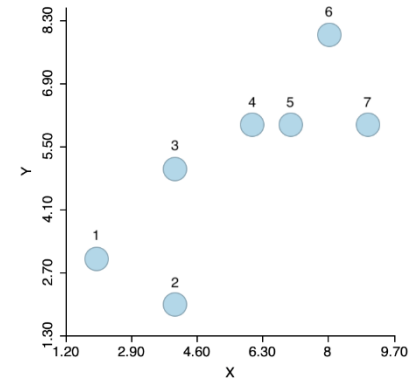
Run Close





# Hierarchické shlukování

- Např.: metoda nejbližšího souseda, Wardova metoda, ...
- **aglomerační** = opakované spojování dvou shluků až do jednoho počínaje jednotlivými objekty jako jednoprvkovými shluky
- **divizivní** = opakované rozdělování nějakého shluku až na jednotlivé prvky počínaje jedním shlukem se všemi objekty
- Grafické zobrazení: **dendrogramu** = stromový diagram
- Vhodné pro aplikace vyžadující hierarchii shluků, např. taxonomie tříd objektů



K Means  
 K Medians  
 K Medoids  
 Spectral

**Hierarchical**

Input:

Select Variables

COUNT  
 AVE\_ID\_  
 dept  
 Crm\_prs  
 Crm\_prp  
 Litercy  
 Donatns  
 Infants  
 Suicids  
 MainCty  
 Wealth  
 Commerc

Use geometric centroids Auto Weighting  
 Weighting: 0 1 1

Select Spatial Weights:

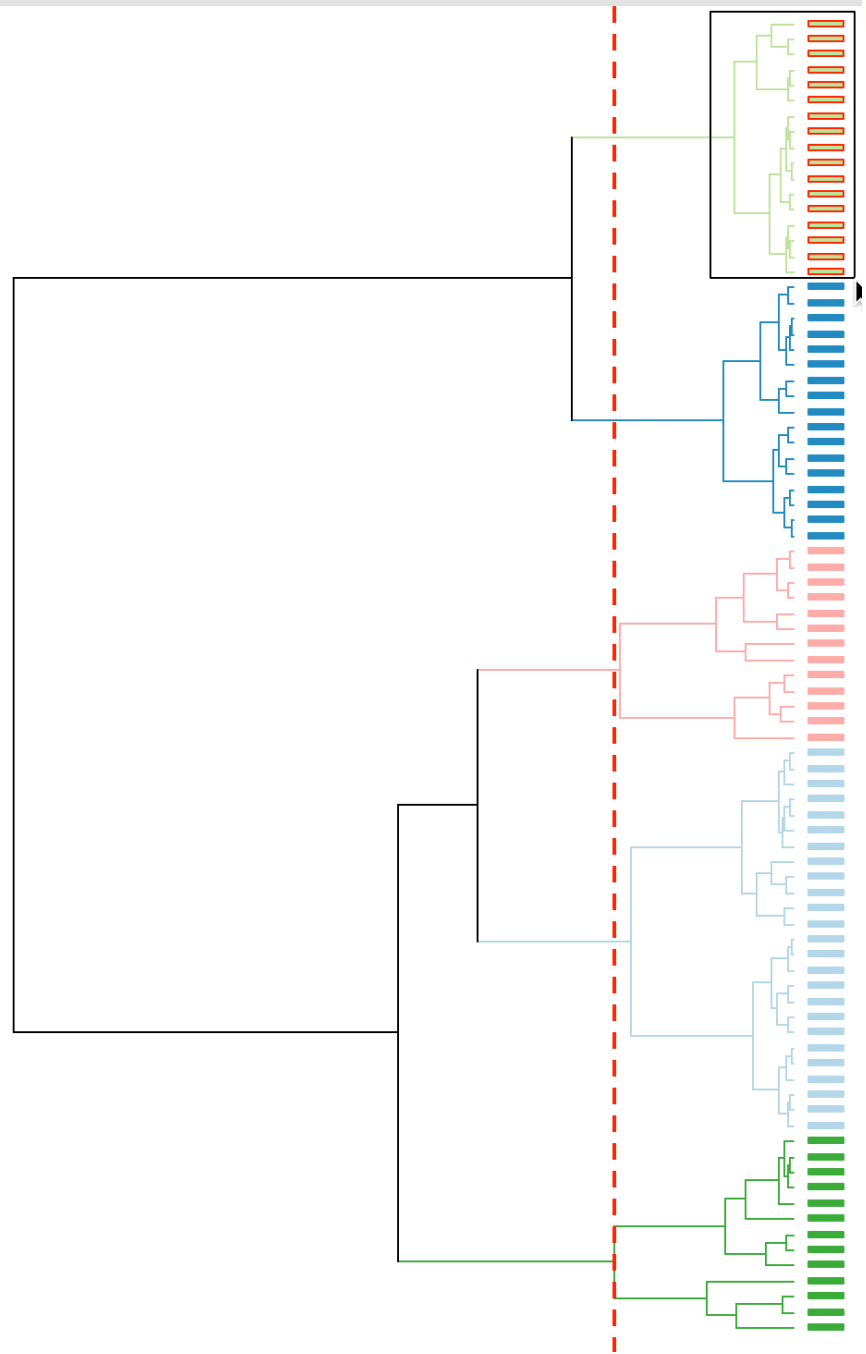
Parameters:

Number of Clusters: 2  
 Transformation: Standardize (Z)  
 Method: Ward's-linkage  
 Distance Function: Euclidean

Output:

Save Cluster in Field: CL

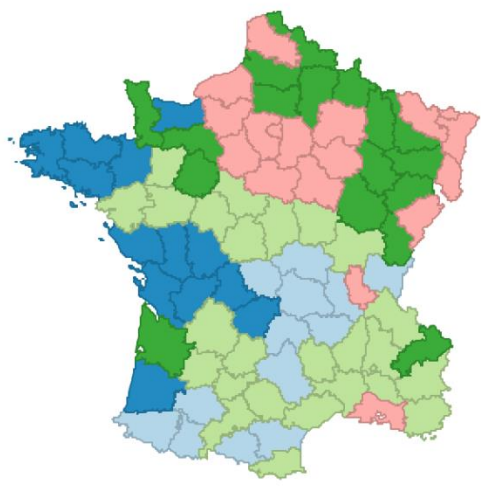
Dendrogram Summary



Hierarchical Cluster Map (5 clusters)

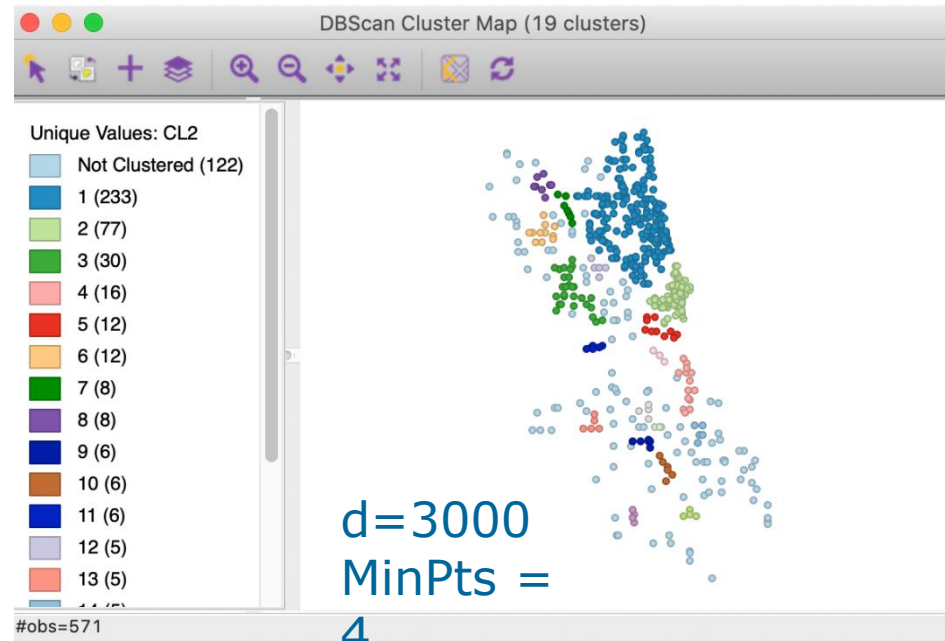
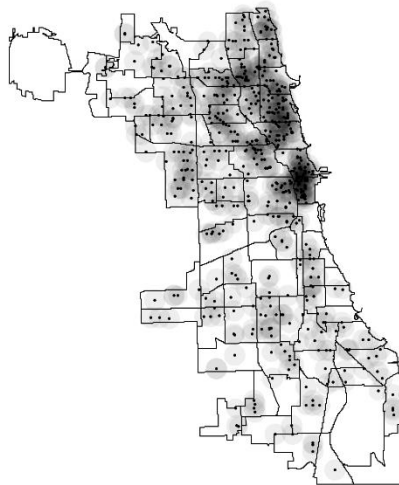


- Unique Values: CLh5
- 1 (13)
  - 2 (13)
  - 3 (25)
  - 4 (17)
  - 5 (17)



# DBSCAN

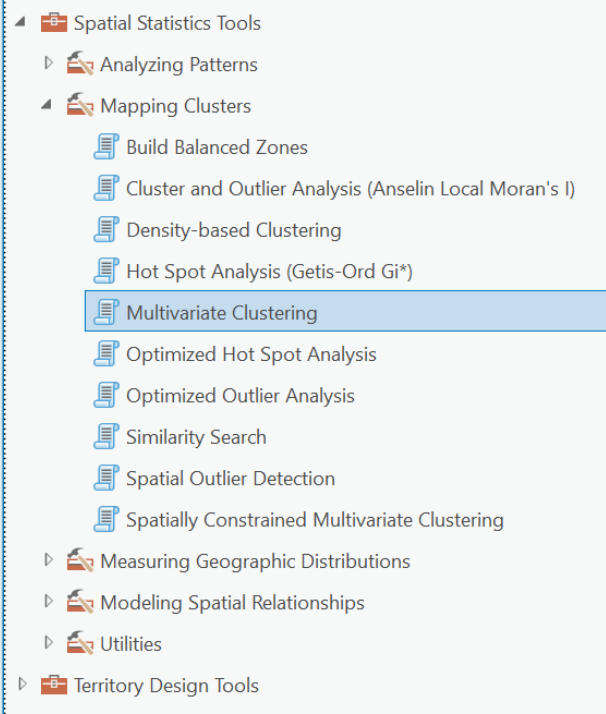
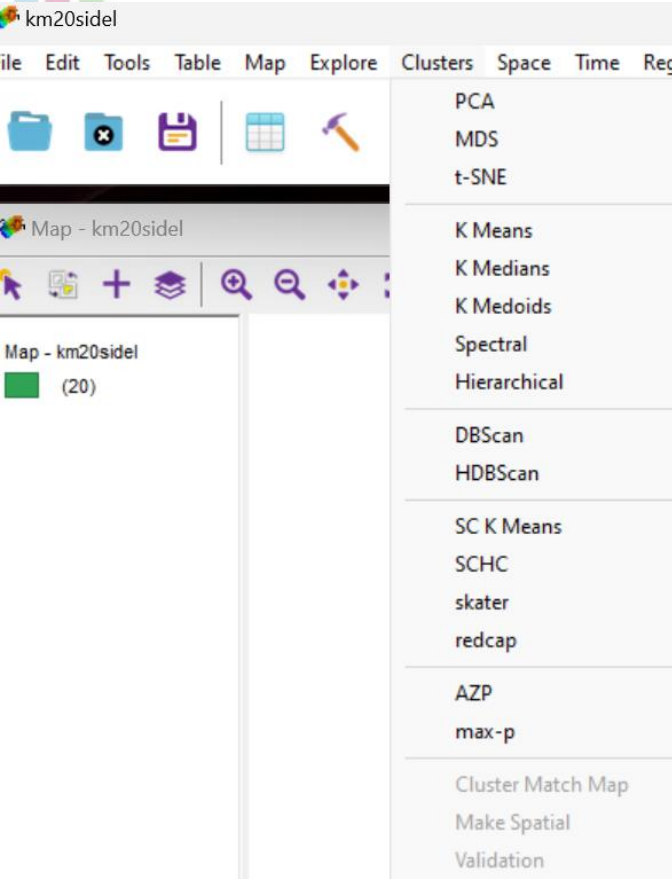
- „Density-based spatial clustering of applications with noise“
- Vychází z hustoty definované pro blízké okolí každého objektu, z dosažitelnosti objektů zjištěné na základě této hustoty a propojenosti dvou objektů ověřené pomocí dosažitelnosti vybraných objektů
- Není založen na vzdálenostech mezi objekty, a tím umožňuje nacházet shluky obecně libovolného tvaru (i shluky uvnitř jiného shluku)
- Nevýhodou je nutnost zadat parametry hustoty, nebo minimální počet prvků ve shluku.



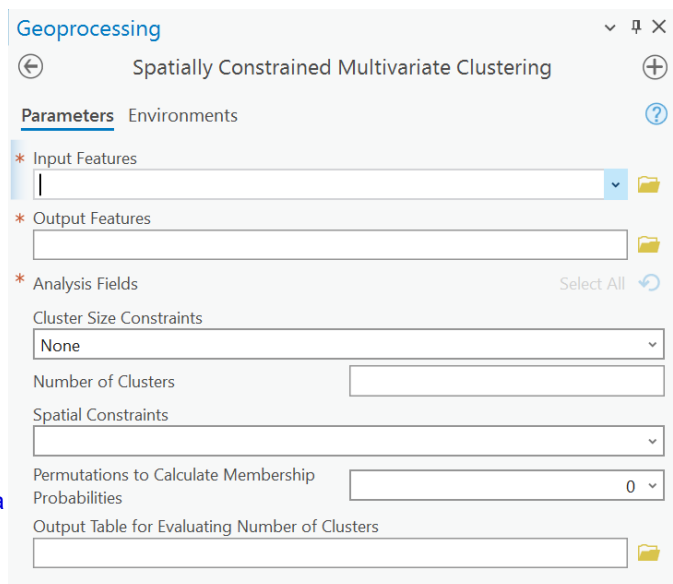


# SOFTWARE

- GeoDa
- QGIS
- ArcGIS Pro



- Rastrová analýza
- Rastrová analýza terénu
- Rastrové nástroje
- Síť (mesh)
- Síťová analýza
- ▾ Vektorová analýza
  - Analýza nejbližšího souseda
  - DBSCAN shlukování (clustering)
  - Join by lines (hub lines)
  - K-means shlukování (clustering)
  - Průměrná/é souřadnice
  - Překryvová analýza
  - Součet délek linií
  - Spočítat body v polygonu
  - ST-DBSCAN clustering**
  - Σ Statistiky podle kategorií
  - Wypis jedinečných hodnot
  - Vystoupat podél linie
  - Vzdálenost k nejbližšímu rozbočovači (body)
  - Vzdálenost k nejbližšímu rozbočovači (linie k rozbočovači)
  - Vzdálenostní matice
  - Σ Základní statistiky pro pole
- Vektorová geometrie



verzita

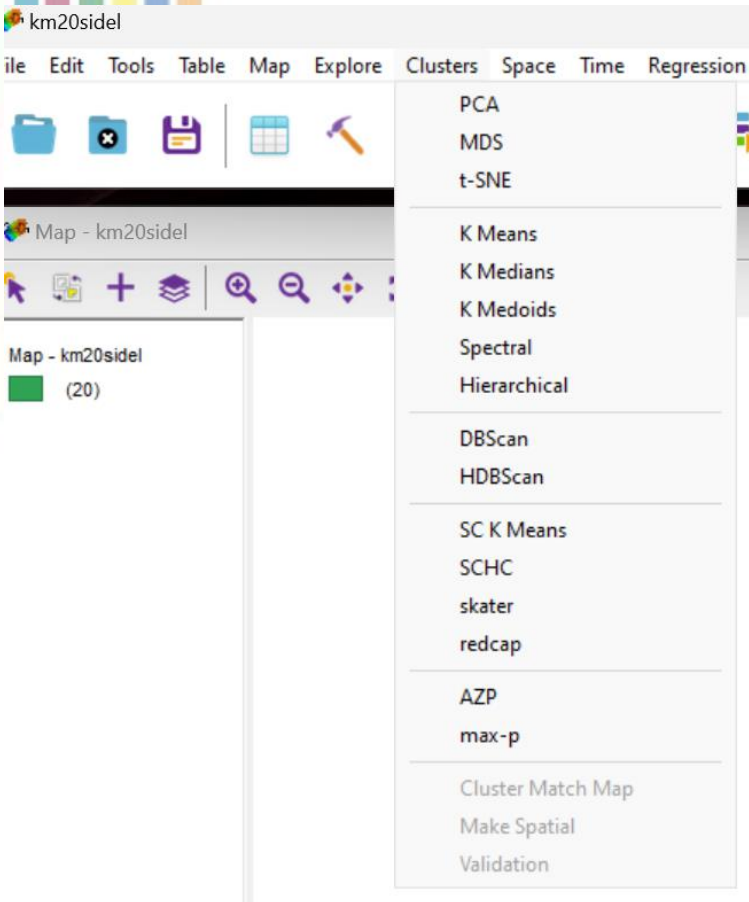


# Další metody multivarianční analýzy

- Analýza hlavních komponent
  - Cílem je redukce původního počtu popisovaných proměnných novými veličinami (umělými), označenými jako komponenty, které shrnují informaci o původních proměnných za cenu minimální ztráty informace.
- Faktorová analýza
  - Cílem je popsat chování množiny cílových proměnných pomocí menšího počtu nových proměnných, označovaných jako faktory
- Diskriminační analýza
  - Slouží k nalezení pravidel resp. funkcí, podle kterých lze rozřadit objekty do jednotlivých známých tříd s využitím hodnot vybraných proměnných (diskriminátory).



# Jak na to ...



- Vyberte si jeden rok...
- Pokuste se identifikovat shluky městských částí s podobnými vlastnostmi
- Vyzkoušejte různé způsoby shlukování a porovnejte je mezi sebou



# A co ještě ...

- Podobně postupujte také při **analýze prostorového uspořádání**

