

Analýza časovo-priestorovej rytmicity obyvateľstva v Bratislavskom kraji



Autor: Bc. Patrik Bučko
Ročník: 2 mZ-RG

Úvod do problematiky

- Štúdium časovo-priestorovej rytmicity patrí pod disciplínu geografie tzv. **Time Geography**.
- „Ústrednou témou výskumu Time Geography je **rozpoznanie času a priestoru ako jednoty**“ (Ellegård, 2018).
- Kľúčovým konceptom Time Geography je teda štúdium geografických javov v **časopriestorovom kontexte**.

Metódy výskumu

- Výskumná časť je založená na báze spracovania a analýzy dát v **Rstudio**.
- Kľúčom je **vytvoriť optimálne zhluky (klastre) obcí** na základe podobnosti ich denného rytmu obyvateľstva.
- Zhlukovanie je realizované podľa 3 základných typov merania vzdialeností časových radov: **Dynamic Time Warping, Shape-Based Distance, Euclidean**.

Testovanie metód zhlukovania časových radov hodinových prítomností SIM kariet na fiktívnych dátach

Euclidean Distance

- Konzervatívna metóda zhlukovania.
- Jednoducho meria lineárnu vzdialenosť od jednej krivky k druhej krivke a **nepovoľuje žiadne dynamické prispôbenie kriviek**.
- **Berie** do úvahy magnitudy kriviek.

Dynamic Time Warping (DTW)

- Liberálna metóda zhlukovania.
- **Povoľuje stláčanie alebo nafahovanie tvaru krivky** na časovej osi tak, aby sa čo najviac podobala centroidu.
- **Neberie** do úvahy magnitudy kriviek, ak sa hodnoty znorمالizujú.

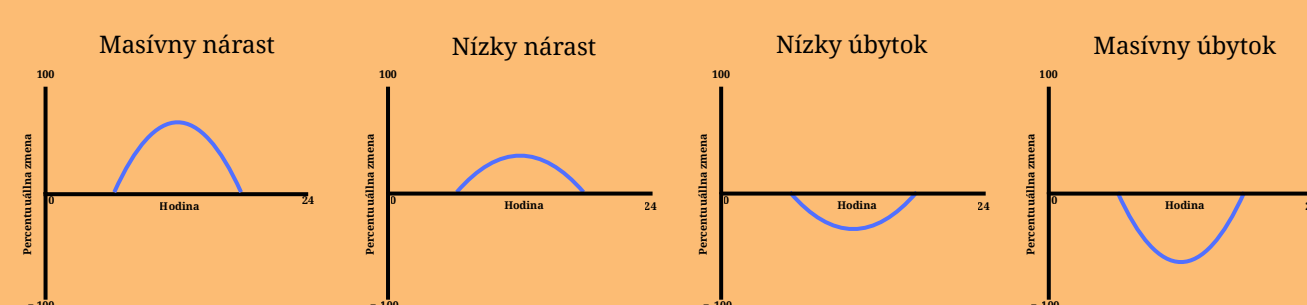
Výsledky klastrovania fiktívnych dát

- **Type = partitional; Centroid = medoid; k = 4**
- Referenčná hodnota je stanovená ako **priemer hodnôt SIM kariet prítomných v obci medzi hodinami 0 až 4 ráno** pre každú fiktívnu obec samostatne.
- Následne je vypočítaná **percentuálna zmena** hodnôt vzhľadom na referenčnú hodnotu.

Shape-Based Distance (SBD)

- Liberálna metóda zhlukovania.
- Porovnáva morfológické tvary kriviek s centroidom **bez stláčania alebo nafahovania krivky na časovej osi, ale povoľuje posun krivky po časovej osi**, aby sa krivka priblížila k centroidu.
- **Štandardne neberie** do úvahy magnitudy kriviek.

Zjednodušený náčrt typov dát vo fiktívnom datasete



Euclidean Distance



DTW



SBD



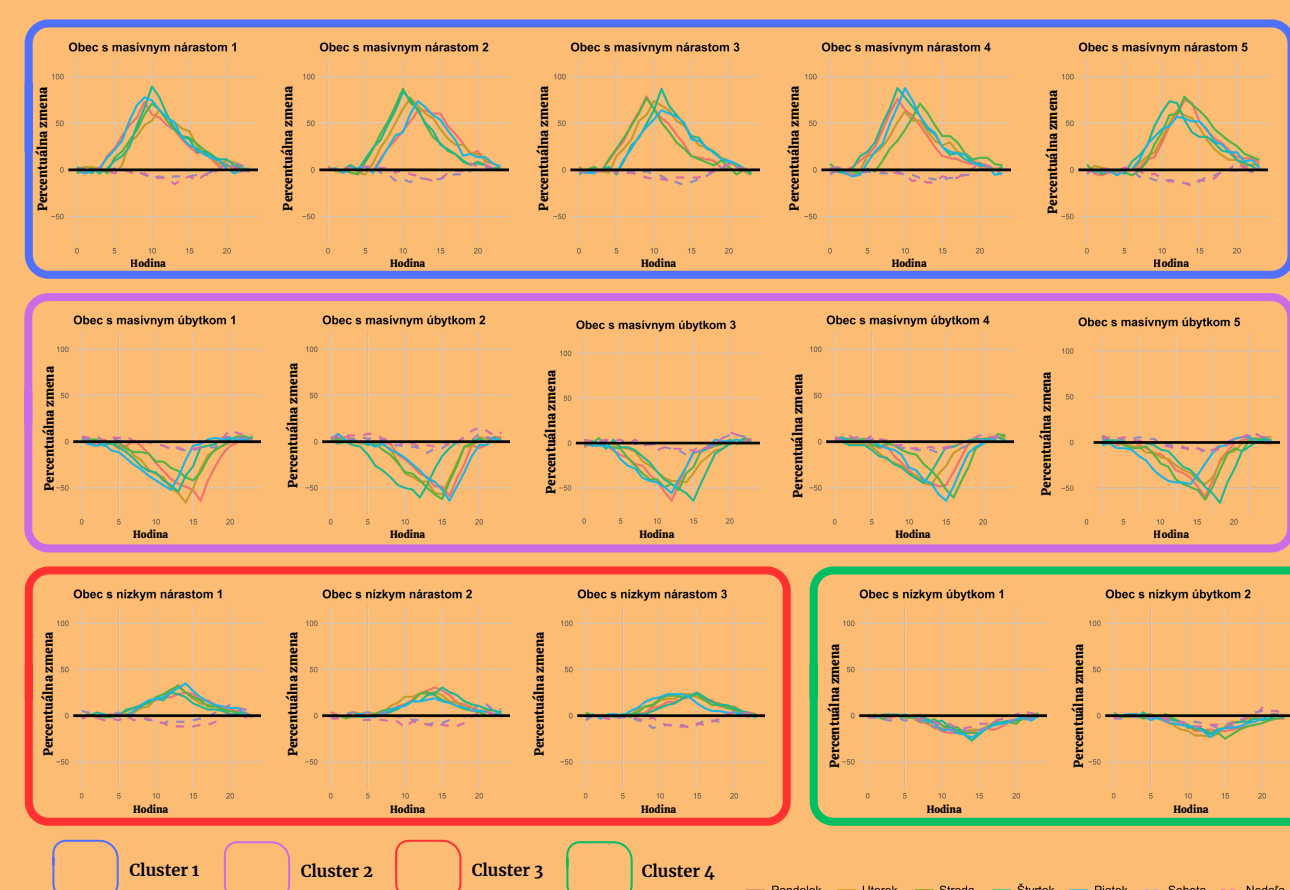
Nevhodnosť týchto 3 metód klastrovania pre tento typ dát. Potreba zvolenia alternatívnej metódy hybridného klastrovania.

- **Výhoda hybridného klastrovania** - berie do úvahy tvar kriviek aj magnitudu kriviek.
- Magnitúda percentuálnej zmeny odráža dynamiku objemu populácie. **Ak sa v našom prípade ignoruje, obce s výrazne odlišnými magnitudami sa môžu priradiť do rovnakého klastra.**
- Čím vyšší je rozdiel v magnitúde medzi obcami, tým väčšiu vzdialenosť prisudzuje medzi krivkami daných obcí.

Type - partitional, Centroid - medoid, k = 4

Výsledok

Euclidean, DTW aj SBD identicky priradili obce do klastrov



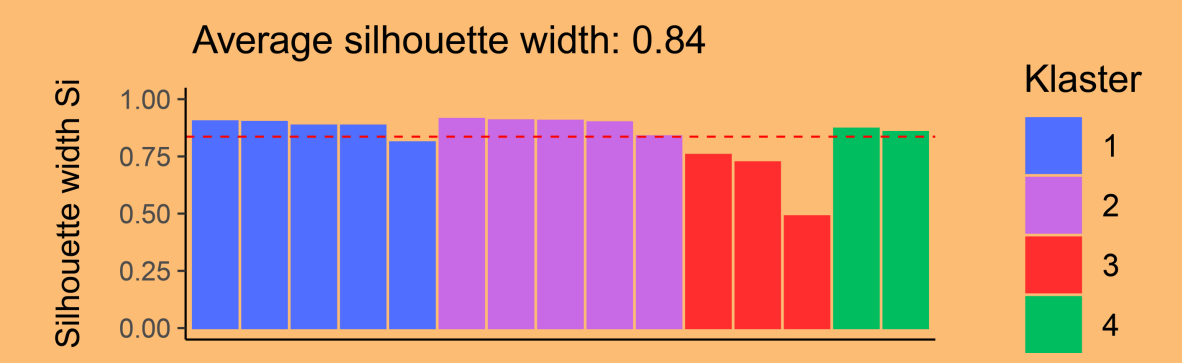
Časť kódu v Rstudio zodpovedná za **zacomponovanie magnitudy** do procesu klastrovania:

```
# Magnitude penalty
feat_a <- magnitude_features[magnitude_features$City == a_name, ]
feat_b <- magnitude_features[magnitude_features$City == b_name, ]

mag_penalty <- abs(feat_a$mean_value - feat_b$mean_value) +
abs(feat_a$amplitude - feat_b$amplitude)

return(sbd_dist + 0.3 * mag_penalty)
```

Silhouette Plot (SBD) - čím bližšie sú hodnoty k 1, tým ideálnejšie je priradenie do klastra; čím bližšie k -1, tým horšie



Aplikácia na reálne dáta

Hybridné klastrovanie na reálnych dátach nefungovalo

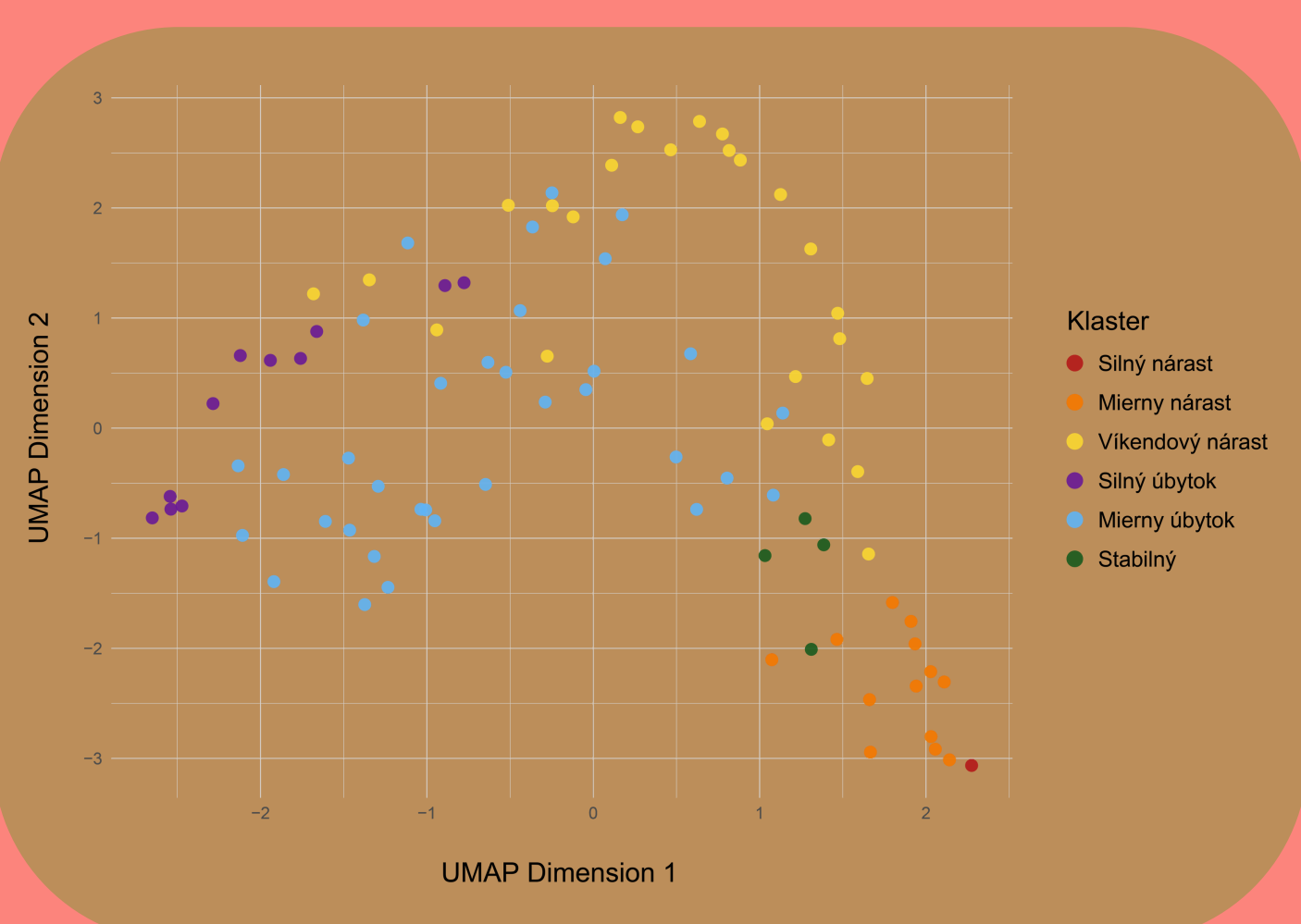
- V reálnych dátach je **príliš veľa variácií a nestability** (nie všetky obce majú ideálne tvary kriviek ako v prípade fiktívnych dát).
- Dáta s ideálnymi tvarmi kriviek, ako v prípade fiktívneho datasetu, sa klastrujú relatívne ľahko. Takéto dáta však častokrát **nereflektujú stav v reálnom svete**.
- **Potreba ďalšieho nastavenia parametrov** pre vytvorenie ideálnych klastrov obcí (iba tvar, vzdialenosť a magnitúda kriviek nestačí).

Aplikácia dvojfázového hybridného klastrovania

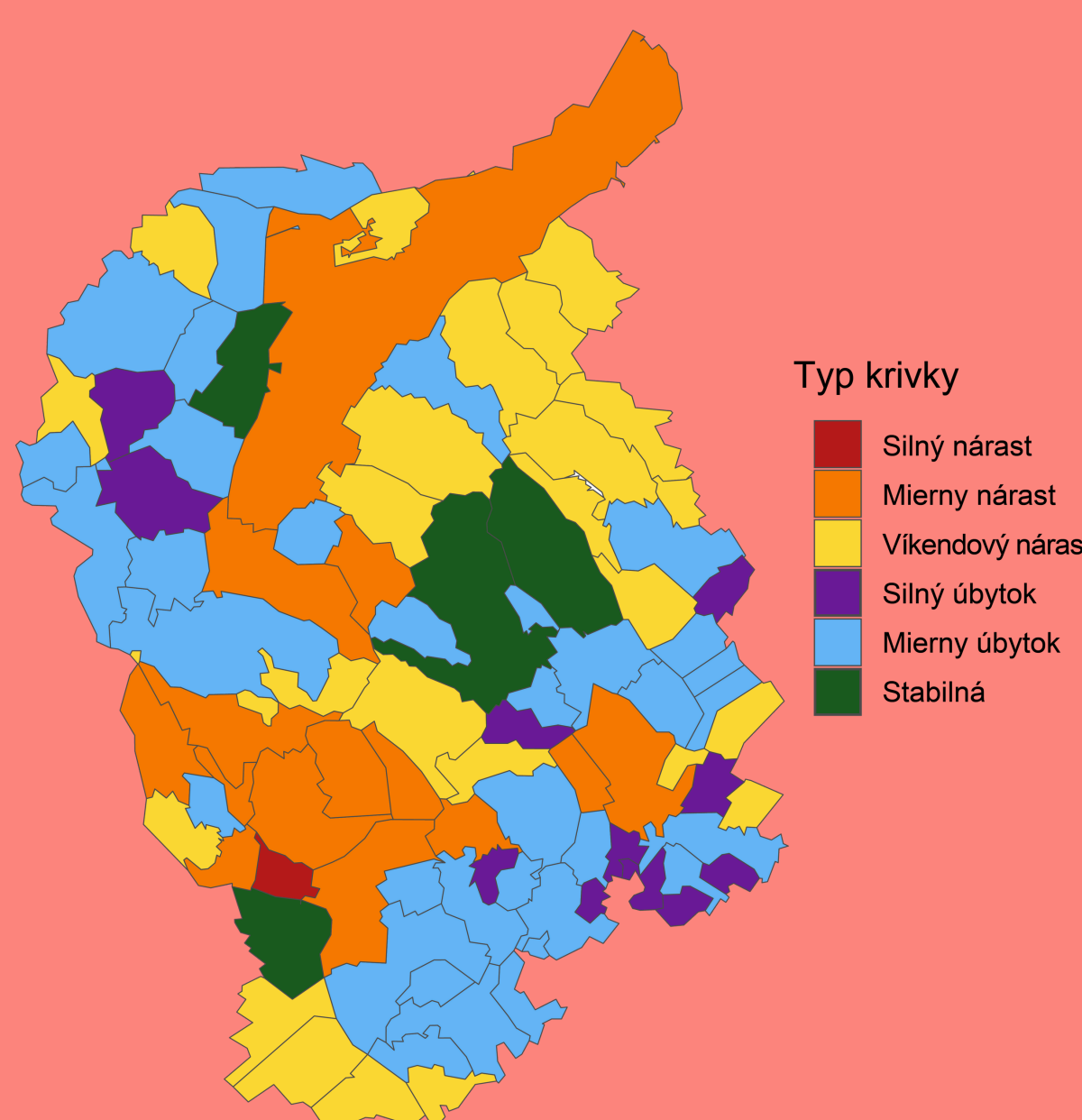
1. **Fáza - výpočet indikátorov:**
 - **Direction Index** - ukazuje, či má mesto tendenciu k rastu alebo poklesu v porovnaní s východiskovou hodnotou.
 - **Amplitúda** - rozdiel medzi maximom a minimom (meria variabilitu).
 - **Net trend** - rozdiel medzi hodinou 167 a hodinou 0 (zmena na konci týždňa a na začiatku týždňa).
 - **Weekend diff** - rozdiel medzi priemerom za víkend a priemerom za pracovný deň (identifikácia špecifických víkendových vzorcov).

2. **Fáza:**
 - Parametre určia 3 skupiny obcí: 1. Obce s jednoznačne **stúpajúcou krivkou**, 2. s jednoznačne **klesajúcou krivkou**, 3. s jednoznačne **stabilnou krivkou**.
 - SBD/DTW následne vytvorí klastre v rámci určených skupín **separátne, nie v rámci všetkých analyzovaných obcí**.
 - Výsledky stále neboli ideálne, 43 z 88 obcí bolo potrebné **manuálne priradiť do správneho klastra**.

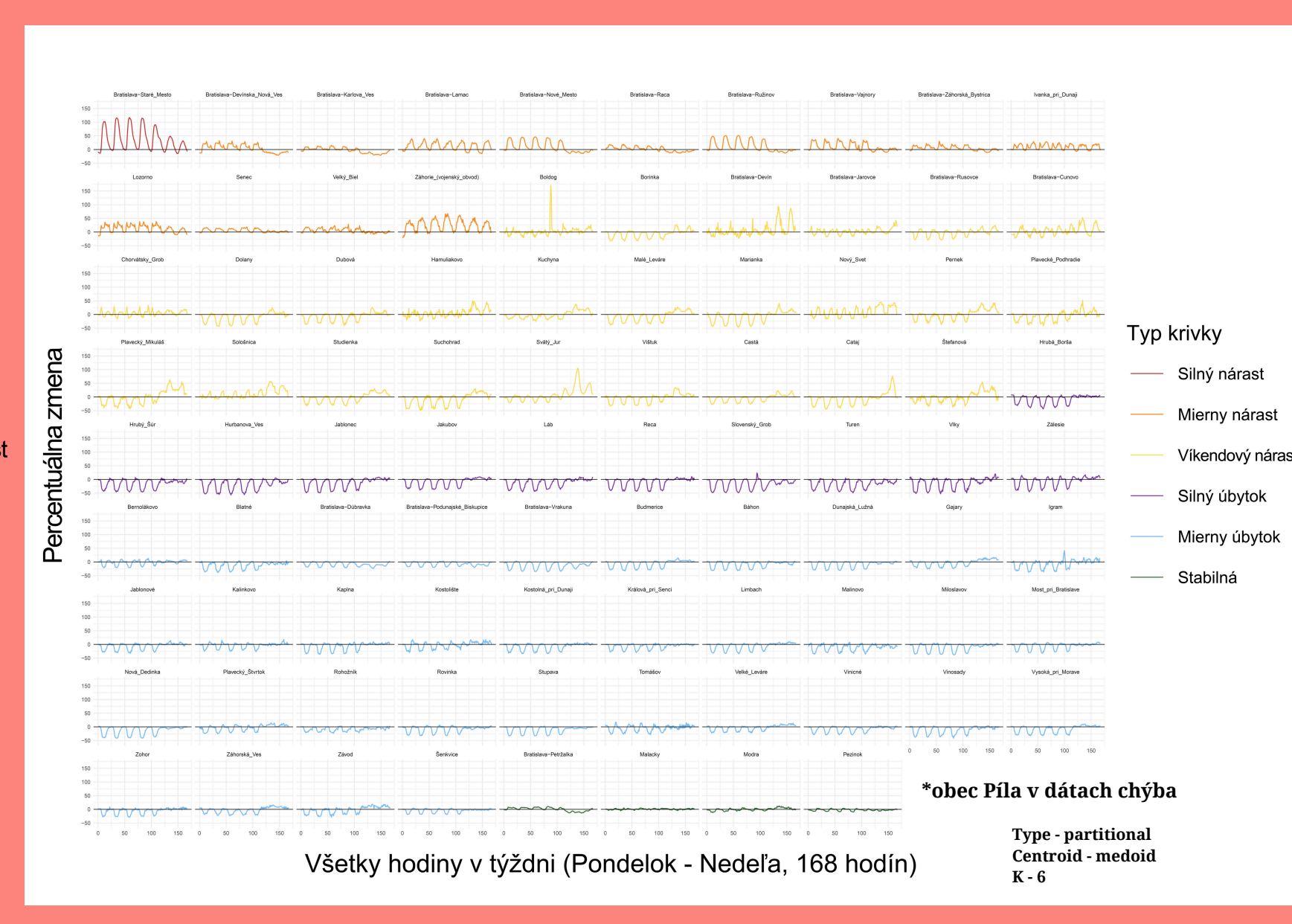
UMAP vizualizácia klastrov (SBD)



Priestorové rozmiestnenie finálnych klastrov (SBD)



Výsledné finálne klastre (SBD)



- Jednotlivé klastre v grafe sú zlúčené pokope, **roztáhanosť niektorých bodov je očakávaná** kvôli charakteru dát z rytmicity reálneho sveta.
- Napriek roztáhanosti sa dá povedať, že **klastre vytvárajú jasné vzorce**, pretože body nie sú chaoticky prepletené medzi sebou, čo by indikovalo nekonzistentné klastre.