

# Hledání optimální polohy stanic a zastávek na tratích regionálního významu



IPR ———  
PRAHA



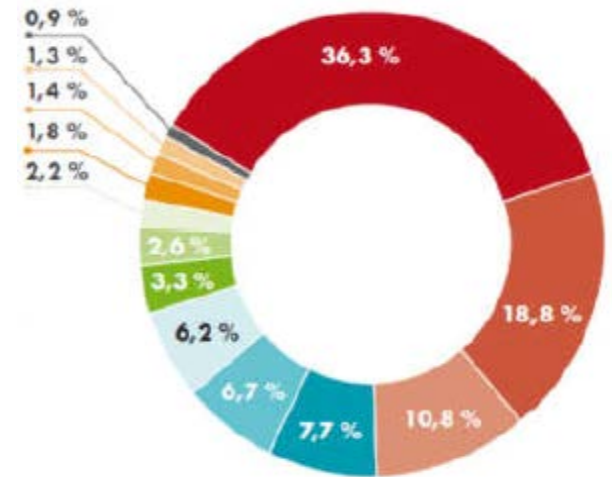
FAKULTA  
DOPRAVNÍ  
ČVUT V PRAZE

# Anotace

1. Dopravní obsluha území tratěmi regionálního významu
2. Cíle výzkumu a algoritmus práce
3. Zdroje dat a databáze relevantních obcí
4. Shluková analýza
5. Regresní analýza a expertní analýza
6. Závěrečná doporučení

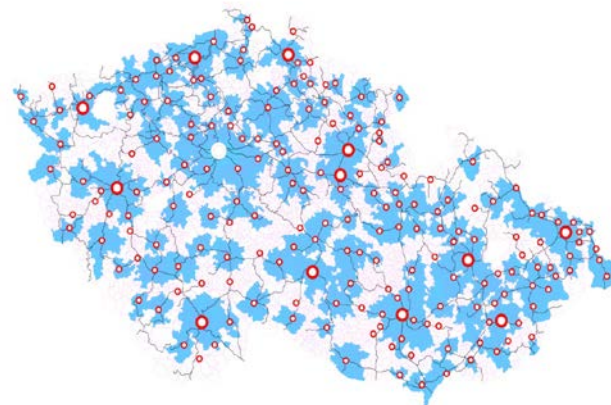
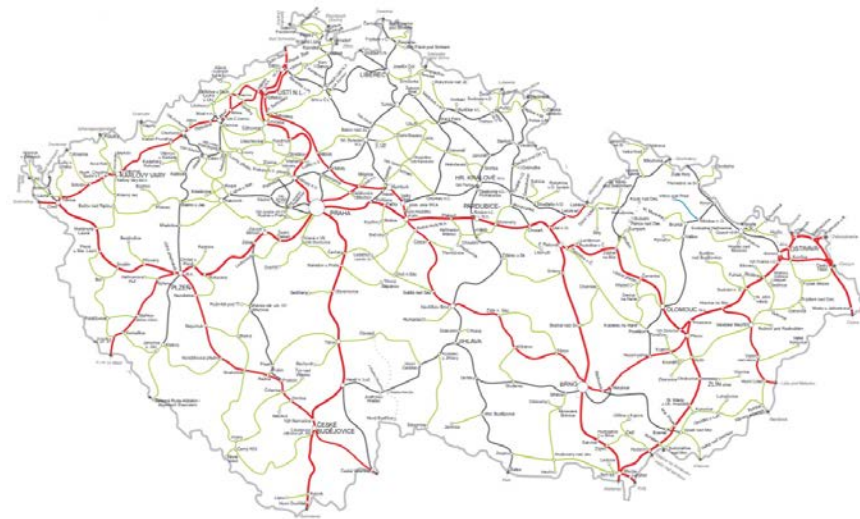
# Problematika dopravní obsluhy území

- ↘ Školy, zaměstnání, úřady, zdravotnictví, kultura
- ↘ Fungující dopravní síť – elementární prvek
- ↘ Nízká dynamika změn rozsahu sítě
- ↘ Udržitelný rozvoj a mobilita
  - ↘ Drážní doprava součástí
  - ↘ Snaha o ovlivňování mobility
  - ↘ Finanční udržitelnost => optimální síť
- ↘ Zastávka/stanice vstupní branou do systému

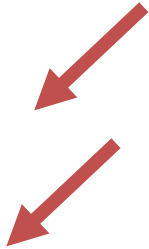


# Vymezení řešených železničních tratí

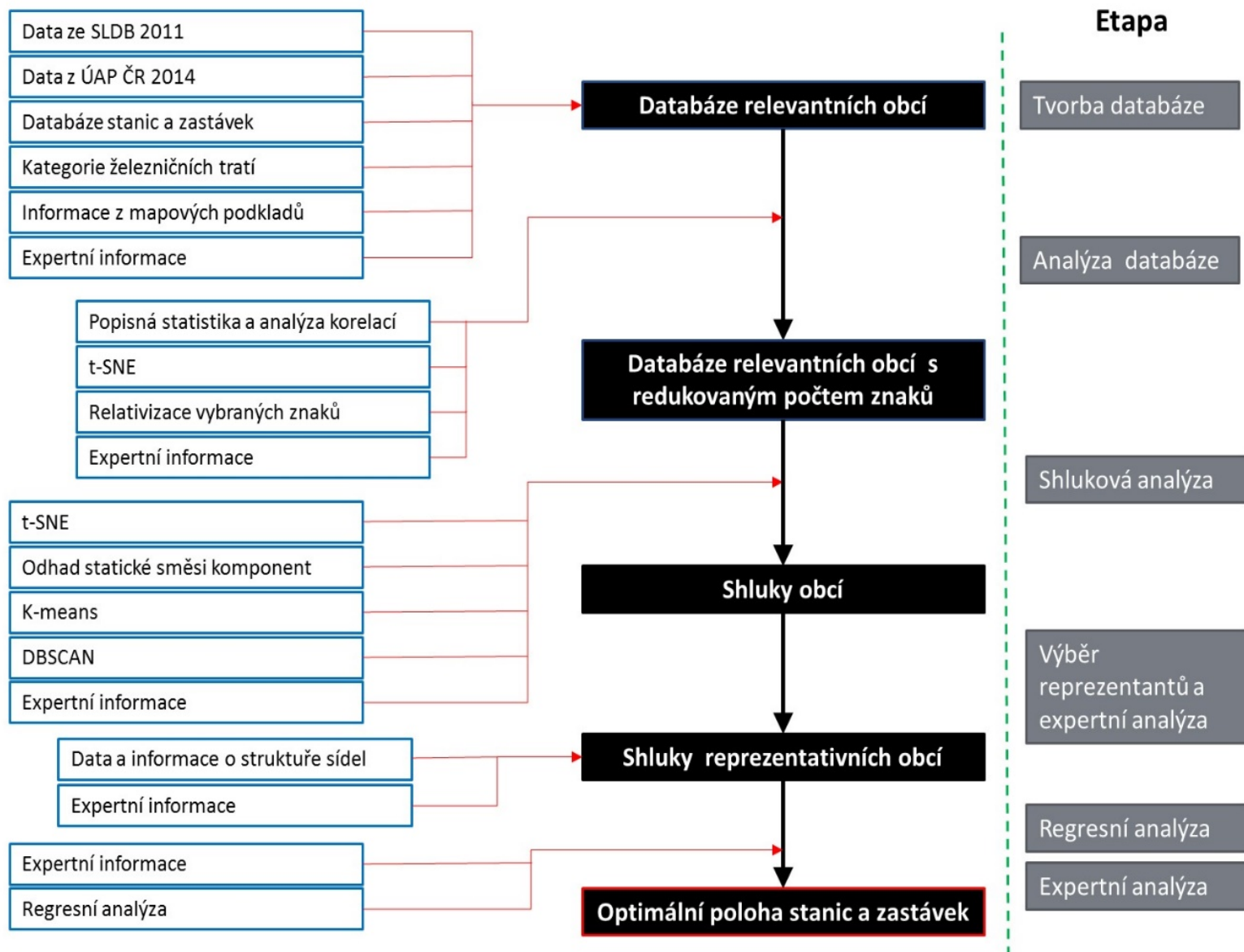
- ↘ Páteřní tratě
- ↘ Tratě regionálního významu
  - ↘ 72% všech tratí
  - ↘ Mimo síť TEN-T
  - ↘ Regionální a některé celostátní tratě
  - ↘ Obsluha menších sídel (bez MHD)
  - ↘ Krajské / lokální / napájecí relace
  - ↘ Historicky primárně pro přepravu komodit



# Cíle výzkumu

- ↘ Systémový nástroj pro optimální polohu žst. / zast. pro maximální využití cestujícími
  - ↘ Hypotéza o ovlivnění volby dopravní módu polohou žst./zast.
  - ↘ Posouzení skutečného potenciálu pro cestující
  
  - ↘ Důležité předpoklady:
    - ↘ Vhodné traťové poměry
    - ↘ Stěžejní relace: bydliště – vzdělání – pracoviště – služby
    - ↘ Záleží na poloze místa zastavení vlaku v sídle
    - ↘ Sídlo má rozčlenitelnou strukturu dle funkčního využití ploch
    - ↘ SLDB: železnice se využívá v obci uvedené obyvatelem
- 

# Metodika hledání optimální polohy žst. / zast.



# Zdroje dat

- Systémové zdroje dat za celou ČR
- Sčítání lidu, domů a bytů 2011
- Územně analytické podklady ČR 2014
- Seznam tarifních bodů dle SŽDC SR 70

## ➤ Mapové služby

➤ Seznam

➤ Google

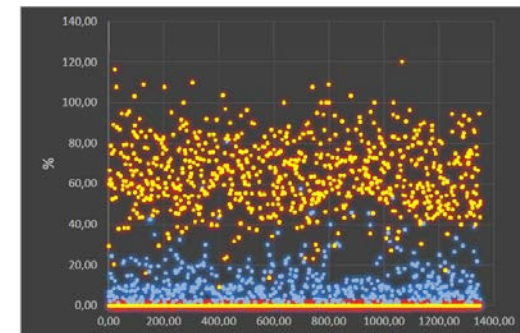
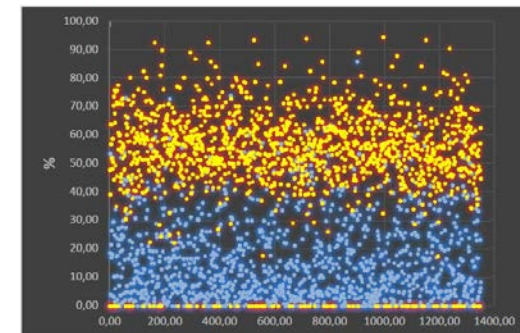
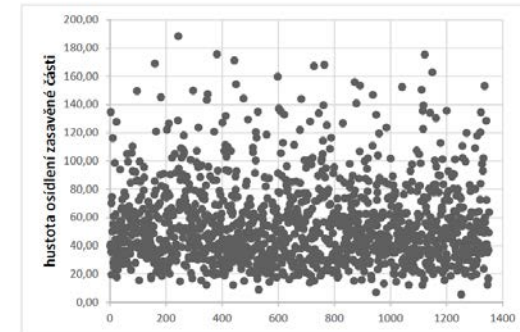
➤ WMS ČÚZK

21. Místo pracoviště nebo školy		na stejné adrese, jaká je v záhlaví formuláře	<input type="radio"/>	jinde v České republice	<input type="radio"/>	uvedte kde ↓
okres						
obec						
č.p./č.o.	/	ulice				
v jiném státě	<input type="radio"/>	uvedte název ↓				zaměstnání bez stálého pracoviště <input type="radio"/>
O každodenní dojízdce nebo docházce do zaměstnání nebo školy uveďte:						
22. Dopravní prostředek Uveďte dopravní prostředek/prostředky, které obvykle používáte při jedné cestě do zaměstnání nebo školy.						
autobus (kromě MHD)	<input type="radio"/>	automobil - spolucestující	<input type="radio"/>	motocykl	<input type="radio"/>	
městská hromadná doprava	<input type="radio"/>	vlak	<input type="radio"/>	jiný	<input type="radio"/>	
automobil - řidič	<input type="radio"/>	kolo	<input type="radio"/>	žádný (pouze pěšky)	<input type="radio"/>	
23. Doba trvání dojížděky/docházky Uveďte, jak dlouho Vám trvá jedna cesta do zaměstnání nebo školy.						
do 14 min.	<input type="radio"/>	15 - 29 min.	<input type="radio"/>	30 - 44 min.	<input type="radio"/>	45 - 59 min. <input type="radio"/>
				60 - 89 min. <input type="radio"/>		90 a více min. <input type="radio"/>
Uveďte, jak často dojíždíte z obce svého bydliště do obce pracoviště:						
24. Frekvence dojížděky do místa pracoviště nebo školy		denně	<input type="radio"/>	týdně	<input type="radio"/>	1 - 2x měsíčně <input type="radio"/>
				jinak	<input type="radio"/>	



# Databáze relevantních obcí

- ↘ Seznam tarifních bodů z 20. 2. 2011 dle SŽDC SR 70
    - ↘ Bez TEN-T, krajských měst, ...
  - ↘ MS Excel – SVYHLEDAT a ruční dotřídění
    - ↘ 1 řádek = 1 obec = 1 ID
    - ↘ Přiřazení znaků ze SLDB, ÚAP a mapových podkladů
      - ↘ Demografické charakteristiky
      - ↘ Územní charakteristiky
      - ↘ Charakteristiky dojížděky a vyjížděky
    - ↘ Výsledná matice 1350 x 109
    - ↘ Příliš velký počet znaků
    - ↘ Nulové hodnoty znaků
- => nutná redukce





# Databáze relevantních obcí – redukce znaků

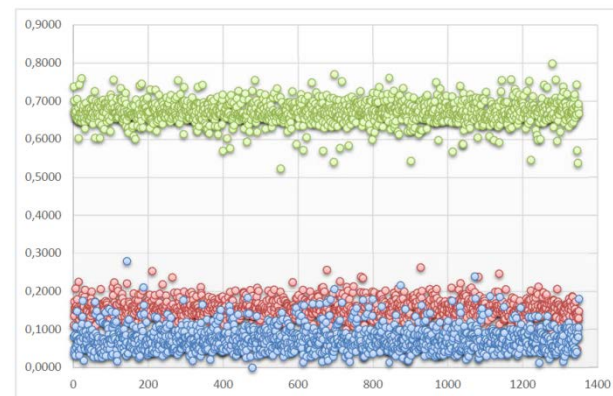
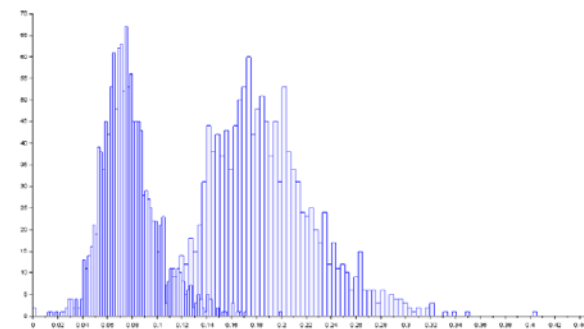
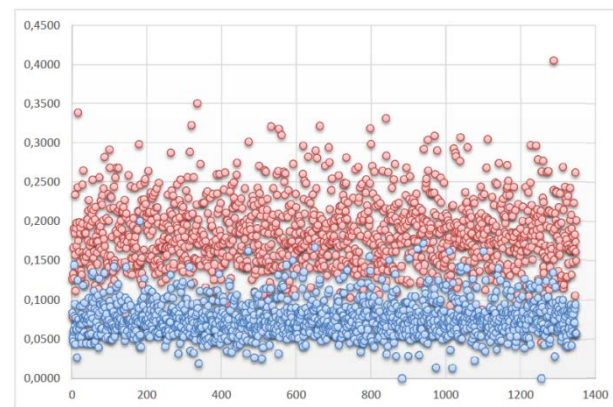
- ↘ t-sne
- ↘ Pearsonův korelační koeficient (109 x 109, standardizovaná data)
  - ↘ Silná až velmi silná korelace ve znacích obsahující počet obyvatel a dojížděku s vyjížděkou
  - ↘ => vybrat reprezentanty korelovaných skupin, volit poměrové znaky
- ↘ Expertní posouzení
  - ↘ Zachování 3 oblastí
  - ↘ Zohlednění nulových hodnot ve sloupcích
- ↘ **Výsledná matice 1350 x 16**



$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

# Databáze relevantních obcí – popisná statistika

- ↘ Pro lepší pochopení struktury a potenciálu dat
  - ↘ Hodnoty základních statistických veličin
  - ↘ Bodové grafy a histogramy četnosti
- ↘ Rozptyl zvláště patrný u znaků
  - ↘ Hustota osídlení pouze v zastavěných částech
  - ↘ Vyjíždka
- ↘ Rizikové znaky s binárními daty
  - ↘ Testováno na SVD a PCA



# Shluková analýza

- ↘ Vícerozměrná statistická metoda pro klasifikaci objektů na základě podobnosti
  - ↘ t-SNE, odhad statické směsi komponent, DBSCAN, K-means
- ↘ Předpoklad: Obce tvoří shluky na základě svého dojížděkového a vyjížděkového potenciálu
- ↘ Míru obsluhy lze optimalizovat (interval, kapacita vozidla)
- ↘ Systémový nástroj – nižší složitost úlohy při respektování rozdílů v potenciálu
- ↘ Problematické nulové hodnoty některých znaků
- ↘ Shluk => statistický výběr

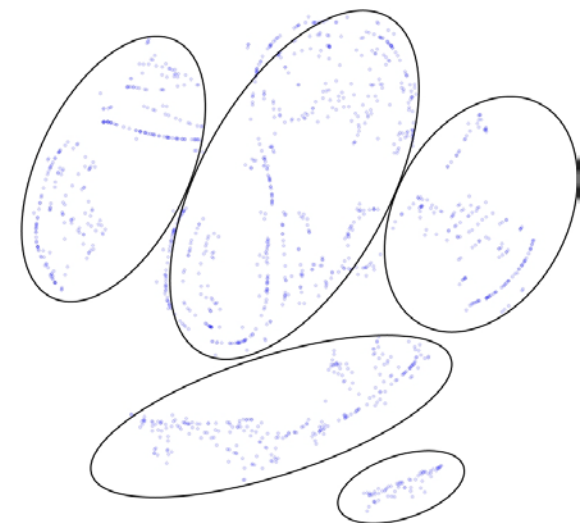
# t-SNE

- ↘ Základní vhled do dat a odhad počtu shluků (matice  $1350 \times 14/16$ )
- ↘ t-distributed stochastic neighbor embedding
- ↘ Nelineární redukce dimenzionality
- ↘ Podobné objekty jsou modelovány blízkými body v 2-3 rozměrném prostoru (Kullback-Leiblerova divergence)
- ↘ Využívá Euklidovskou vzdálenost

t-SNE



t-SNE

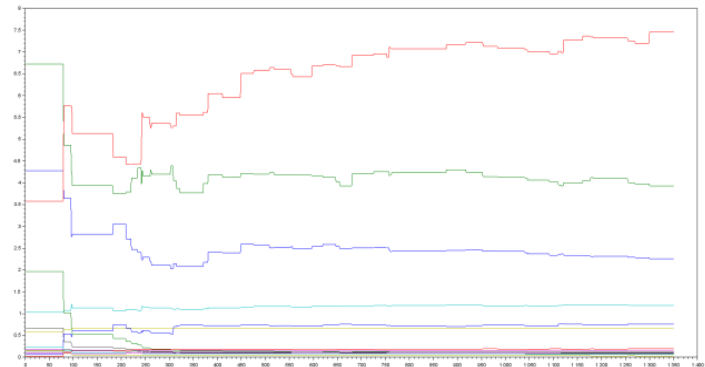


# Odhad statické směsi komponent

- ↘ Transparentní klasifikace naměřených dat (odhad jízdního módu jedoucího automobilu z dat o jízdě)
- ↘ Shluk = komponenta
- ↘ Model tvořen množinou  $n_c$  komponent a diskrétní veličinou  $c_t$ 
$$f(d_t|c, \Theta_c, \psi_t), c \in \{1, 2, \dots, n_c\} = c^*$$
- ↘ Využívá Bayesovské statistiky (vazba na apriorní data) a Gaussova rozdělení hustot pravděpodobnosti vektoru vah jednotlivých komponent
- ↘ Hledání center 5 komponent pro disjunktní dvojice znaků ve 3 iteracích
- ↘ Klasifikace objektů (obcí) ve 14-ti rozměrném prostoru
- ↘ Počet apriorních dat  $n_i = 10$

# Odhad statické směsi komponent

- ↘ Výsledky – počet obcí ve shlucích:
  - ↘ **856, 271, 147, 55, 21**

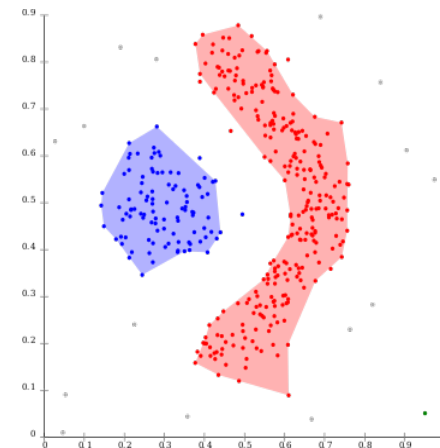


# DBSCAN

- ↳ Density-based spatial clustering of applications with noise
- ↳ Shluky jsou oblasti v prostoru s větší hustotou objektů
- ↳ Každý objekt má definované své vlastní sousedství  $\epsilon$

- ↳ Počáteční podmínky

- ↳ MinPts = 5, zkoušeny i hodnoty větší
- ↳ Hodnoty  $\epsilon$  testovány pro počet shluků 5
- ↳ Testován také počet znaků 8 pro vyšší rovnoměrnost ve shlucích



- ↳ Výsledky – počet obcí ve shlucích:

- ↳ **183, 1139, 16, 14, 5**



# K - means

- ↘ Metoda k-průměrů (k-centroidů s euklidovskou vzdáleností)
- ↘ Iterativní optimalizační metoda založená na Euklidovské vzdálenosti a určení počátečních centroidů

- ↘ Počáteční podmínky

$$D_{euklid.}(x_i, x_j) = \sqrt{\sum_{l=1}^m (x_{il} - x_{jl})^2} = \|x_i - x_j\|$$

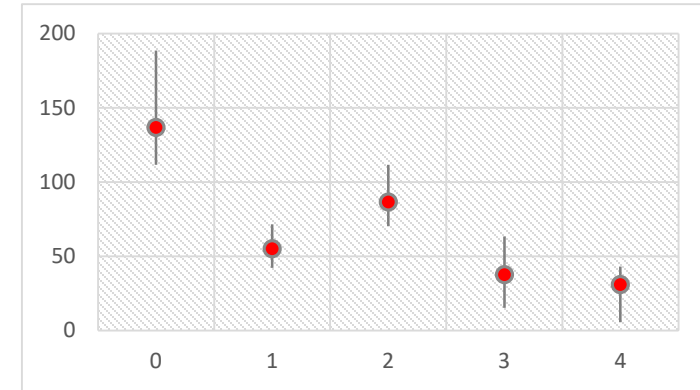
- ↘ Počet shluků = 5, testováno také 3, 5, 6, 7 a 10 shluků
- ↘ Testován také počet znaků 8 pro vyšší rovnoměrnost ve shlucích
- ↘ Průběžná kontrola pomocí datových pruhů MS Excel
- ↘ Výsledky – počet obcí ve shlucích:
  - ↘ **69, 477, 253, 28, 523**

# Shluky obcí

↘ Porovnání metod pomocí datových pruhů a hodnot statistických veličin

↘ 8 znaků s nestandardizovanými daty

↘ Reálná interpretace shluků



↘ **Nejlépe vyhodnocena metoda K-means pro 8 znaků**

↘ „Datově konzistentní“

↘ „Nejrovnoměrnější“ rozložení objektů ve shlucích

↘ Metody založené na hustotě nejsou vzhledem k tvaru dat zřejmě vhodné

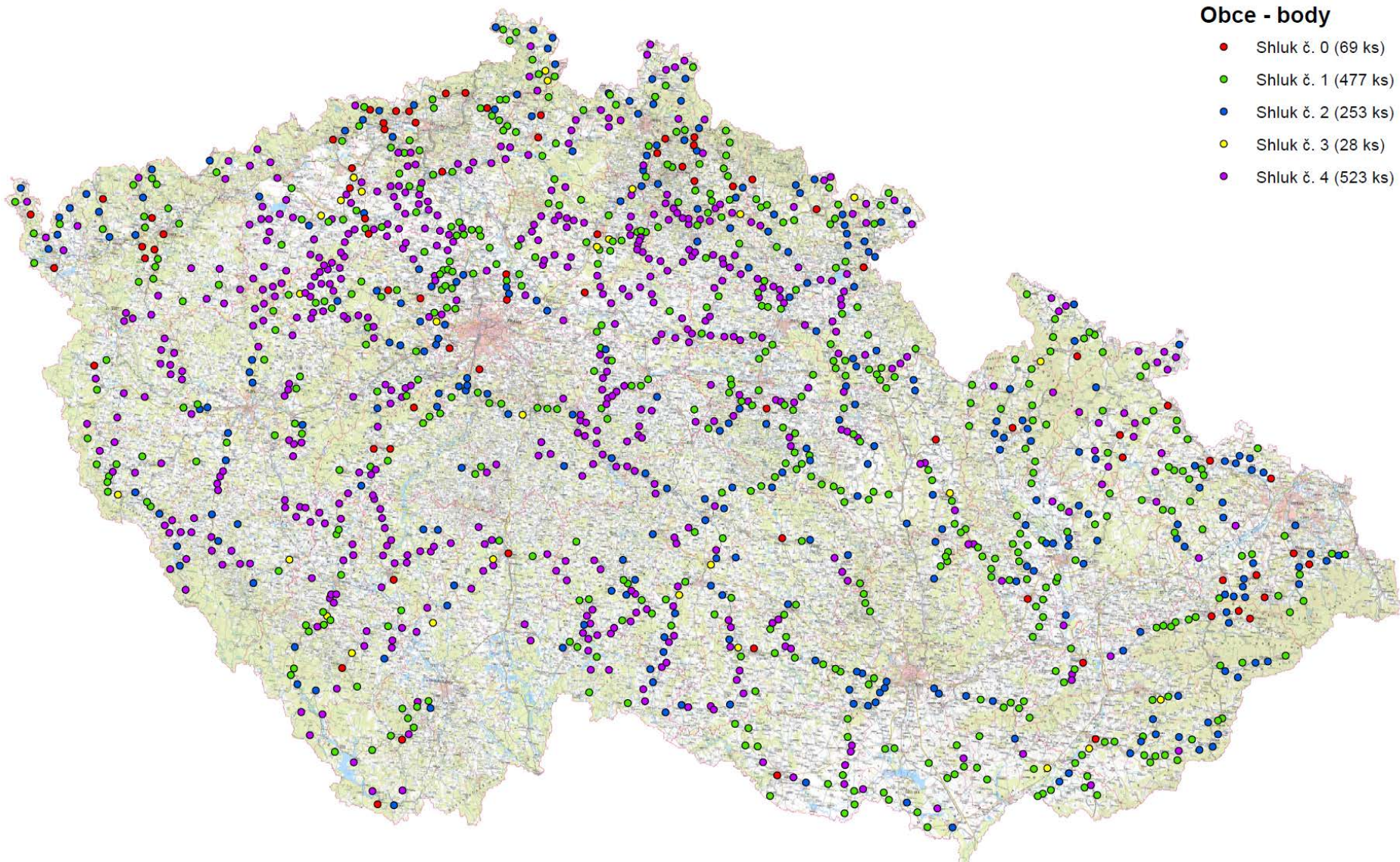
↘ Nutná optimalizace odhadu statické směsi komponent

# Zobrazení shluků obcí na mapě ČR

## Legenda

### Obce - body

- Shluk č. 0 (69 ks)
- Shluk č. 1 (477 ks)
- Shluk č. 2 (253 ks)
- Shluk č. 3 (28 ks)
- Shluk č. 4 (523 ks)



# Shluky reprezentativních obcí a doplnění dat

- ↘ Shluky reprezentativních obcí
  - ↘ Dostupná data o dojížděce a vyjížděce vlakem
  - ↘ Pouze 1 stanice či zastávka na území
- ↘ Výsledný počet 272 reprezentativních obcí ve shlucích:
  - ↘ **19, 117, 88, 10, 38**
- ↘ Doplnění dat z mapových podkladů pro reprezentativní obce
  - ↘ Poloha stanice / zastávky v území
  - ↘ Typ funkční plochy do 500m a 1000m od zastávky
  - ↘ Kvantifikace dat (10 znaků) pro regresní analýzu

# Regresní analýza

- ↘ Vícenásobná lineární regrese se zohledněním náhodné chyby  $E$

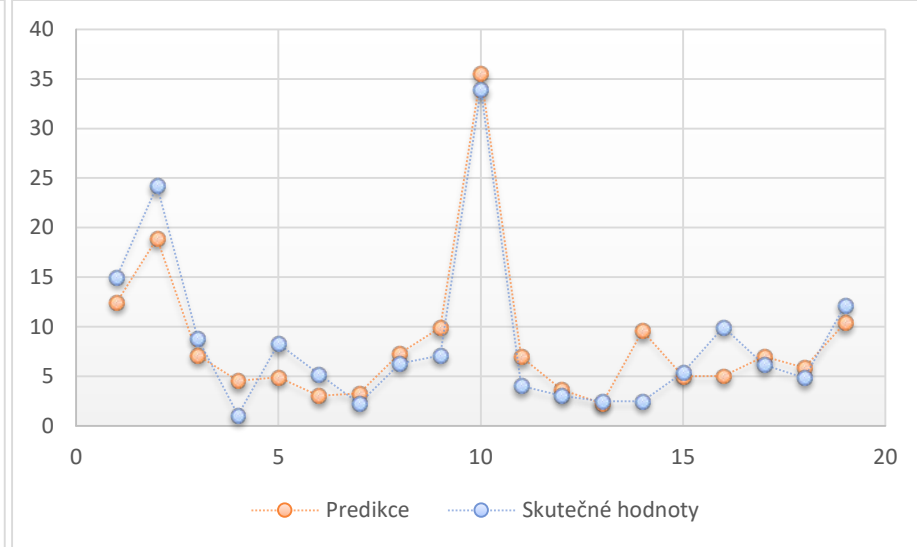
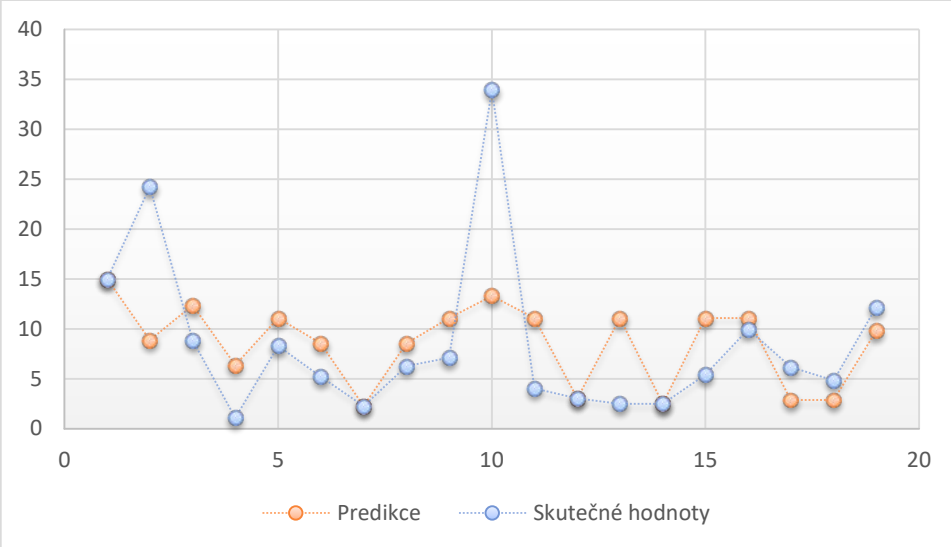
$$Y = k_1x_1 + k_2x_2 + \dots + k_nx_n + q + E$$

- ↘ Testovány i nelineární spojnice trendu dat
- ↘ Regresor = podíl vyjíždějících a dojíždějících vlakem
- ↘ Regresandy = data o funkčním využití území a poloze stanice / zastávky
- ↘ Pro porovnání testovány také tzv. demografické regresandy
- ↘ Testován koeficient determinance a  $p$ -hodnota  $F$ -testu
- ↘ Shluk č. 3 vzhledem k nedostatku dat netestován



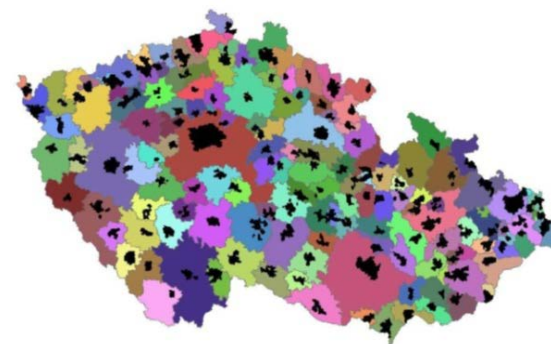
# Regresní analýza - výsledky

Regresandy	Veličina / Shluk	0	1	2	4
Území	Koef. determinance $R^2$	0,242931	0,188578	0,161819	0,415085
	$p$ -hodnota F-testu	0,846350	0,010879	0,160636	0,829693
Demografické	Koef. determinance $R^2$	0,862045	0,148506	0,249070	0,450735
	$p$ -hodnota F-testu	0,021271	0,054834	0,008334	0,046710



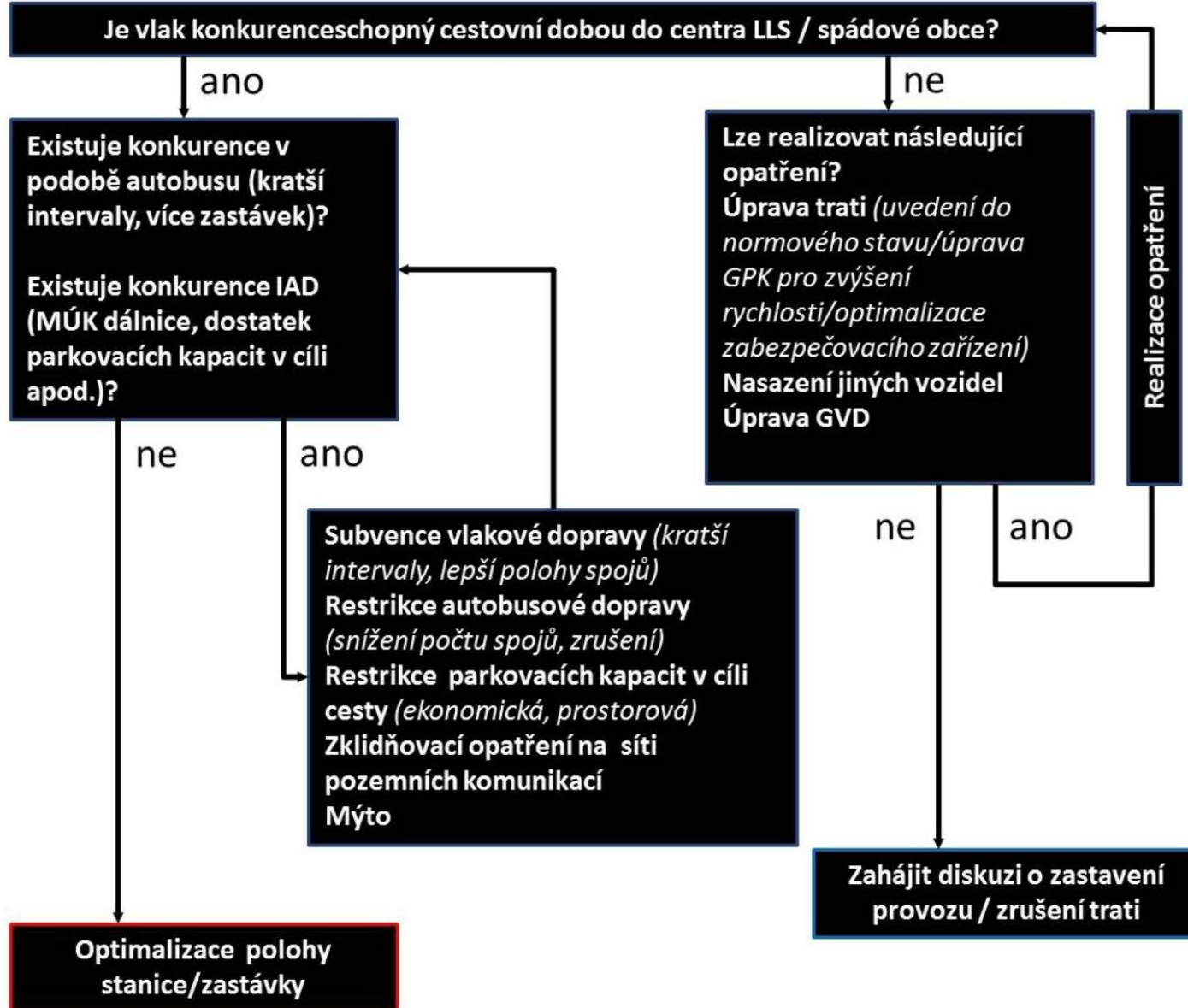
# Expertní analýza

- ↘ Vzhledem ke slabé odezvě regresní analýzy a nutné analýze shluku č. 3
- ↘ Sledování
  - ↘ jízdních dob různými módy do LLC
  - ↘ Polohy stanic a zastávek vlaku a autobusů
  - ↘ Vybavenost obce
  - ↘ Poloha velkých zaměstnavatelů (shluk č.3)
- ↘ Nezanedbatelný vliv na volbu dopravního prostředku
  - ↘ Cestovní doba
  - ↘ Konkurenceschopnost vlakové dopravy
  - ↘ Možnosti parkování v cíli
  - ↘ Navázání na začátky a konce směn (shluk č.3)
  - ↘ Dostatek zastávek u zdrojů (shluk č. 3)



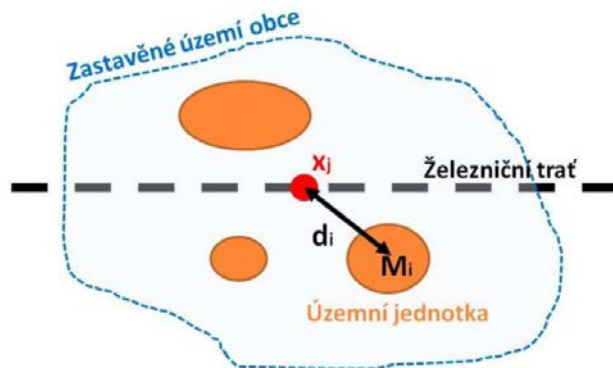


# Závěrečná doporučení



# Závěrečná doporučení

- Poloha stanice či zastávky není sama o sobě dominantním prvkem v rozhodování
- Komplexní problém
- Pro optimalizaci polohy stanice či zastávky konkurenceschopné vlakové dopravy platí:



$$LOC_{OPT} = \max_{j=1,2,\dots,m} \left( \left( \sum_{i=1}^n \frac{M_i}{d_i} \right)_{x_j} \right)$$

- Doplnit data ve SLDB a zjemnit data o členění obcí

# Děkuji za pozornost



IPR  
PRAHA



FAKULTA  
DOPRAVNÍ  
ČVUT V PRAZE

Ing. Václav Novotný, Ph.D.  
Institut plánování a rozvoje hl. m.  
Prahy

Sekce infrastruktury  
Vyšehradská 57  
128 00 Praha 2

E) [novotny.v@ipr.praha.eu](mailto:novotny.v@ipr.praha.eu)

T) +420 236 005 651

M) +420 773 772 172