

Využití dat z mobilních sítí v geodemografickém výzkumu

Petr Mazouch

Katedra ekonomické statistiky

Fakulta informatiky a statistiky VŠE v Praze

Program diskusního večera

- Představení projektu
- Současné problémy s daty z veřejně dostupných zdrojů
- Co jsou vlastně data z mobilních sítí?
- Vztah dat z mobilní sítě a prostorové rozmístění (a pohyb) obyvatel
- Typy uživatelů
- Opakování chování

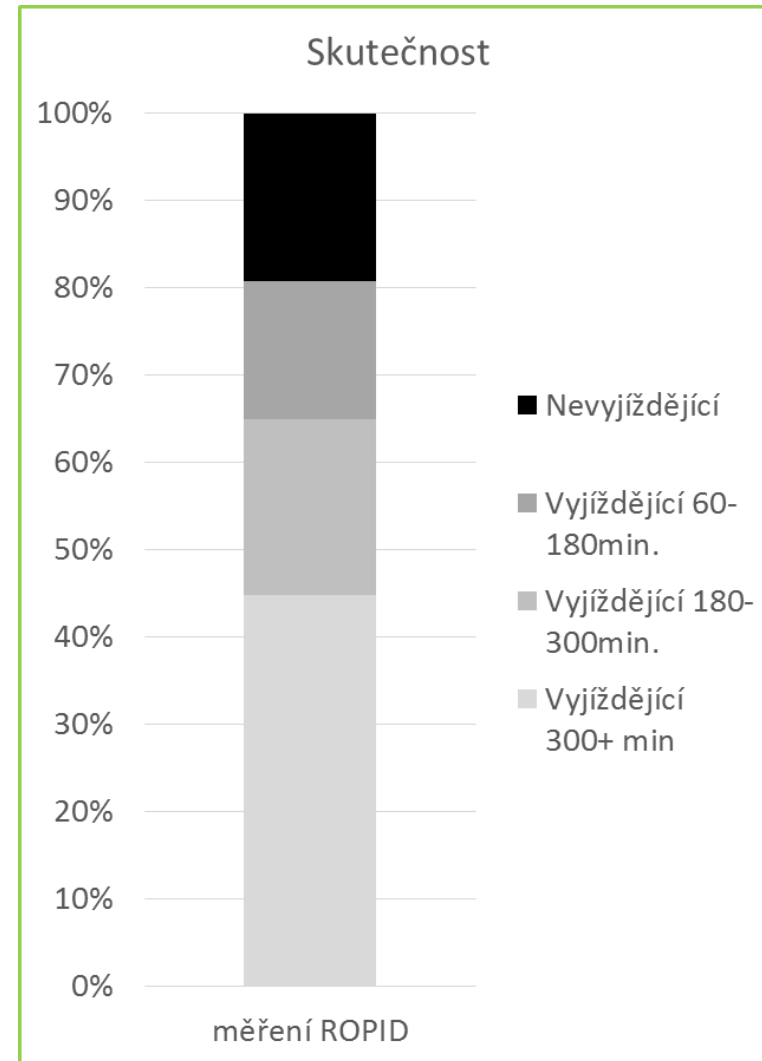
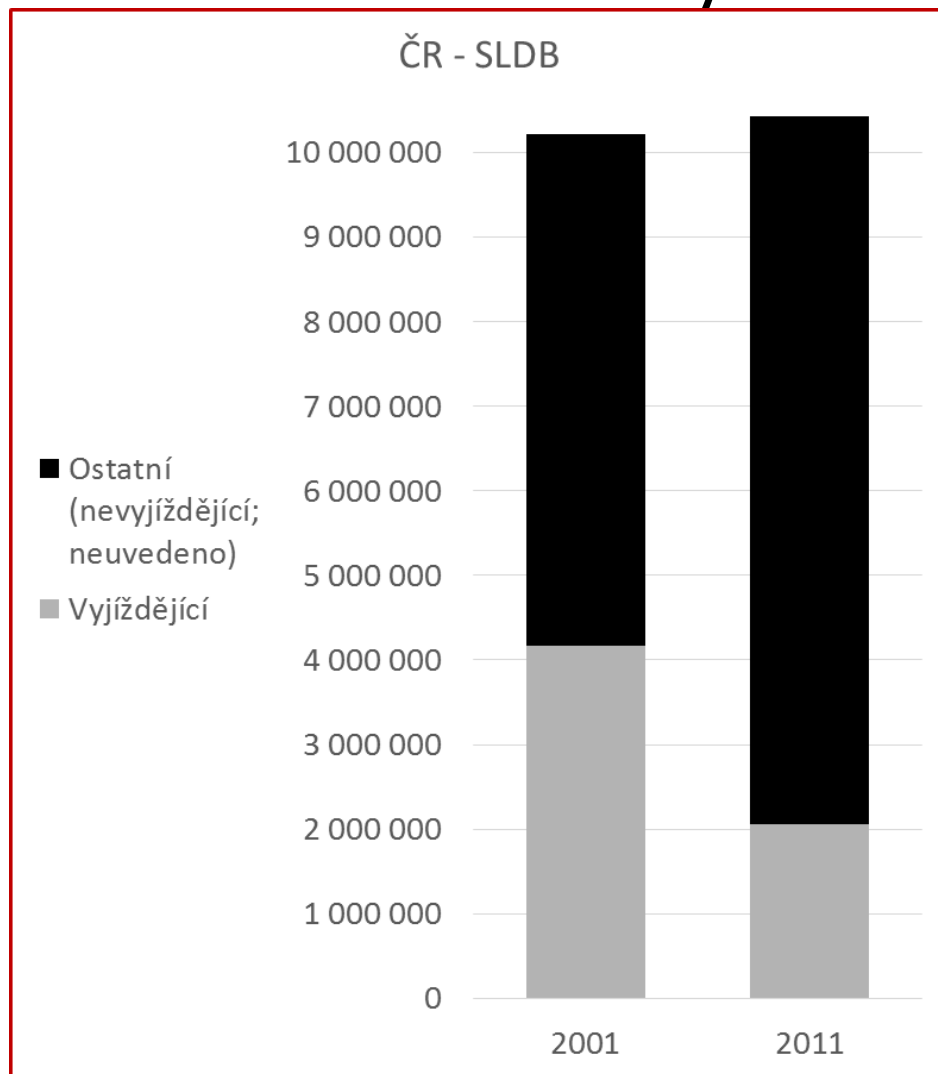
Účastníci projektu

- VŠB
 - VP č. 5, IT4I pro zpracování znalostí, plnění integračního vztahu k záměrům smluvních stran v oblasti big data, návrhy algoritmů pro data mining a jejich implementace s využitím výpočetního výkonu národního superpočítačového centra
- ČVUT
 - FEL ... znalosti v oblasti mobilních technologií a jejich bezpečnosti, sítí nové generace NGN a internetu a správy kmitočtového spektra
 - FD ... znalosti dopravních systémů a matematického modelování
- VŠE
 - znalosti v oblasti informačních systémů, statistických, ekonometrických a demografických metod v hospodářském životě a znalostí problematiky „open data“
- MODATA
 - znalostí ekonomických a sociálních dopadů technologických řešení, strukturování zadání požadavků pro projekty VaVal, testování a odladění výsledků projektů VaVal a řízení zavedení a udržitelnosti výsledků projektů VaVal do praxe

Projektový tým

- VŠB - TECHNICKÁ UNIVERZITA OSTRAVA, pracoviště IT4Innovations Národní superpočítačové centrum; 17. listopadu 15/2172, Ostrava – Poruba, Česká republika
 - Garant: prof. RNDr. Václav Snášel, CSc., (vaclav.snasel@vsb.cz); doc. Ing. Miroslav Vozňák, Ph.D. (miroslav.voznak@vsb.cz)
- České vysoké učení technické v Praze, fakulta elektrotechnická; Technická 2, Praha 6 - Dejvice, Česká republika
 - Garant: prof. Ing. Boris Šimák , CSc. (simak@fel.cvut.cz); Ing. Robert Bestak, Ph.D. (robert.bestak@fel.cvut.cz)
- České vysoké učení technické v Praze, fakulta dopravní; Na Florenci 25, Praha 1, Česká republika
 - Garant: doc. Ing. Ondřej Přibyl, Ph.D.. (pribylo@fd.cvut.cz)
- Vysoká škola ekonomická v Praze, fakulta informatiky a statistiky; nám. W. Churchilla 4, Praha 3, Česká republika
 - Garant: Ing. Petr Mazouch, Ph.D. (mazouch@vse.cz)
- MODATA, a.s., Na Strži 65/1702, 140 62 Praha 4, Česká Republika
 - Garant: Bc. Jiří Hylmar, MBA (jiri.hylmar@modata.cz)

Neexistují informace o denní přepravní poptávce 6 mil. obyvatel



Co nám říkají dostupná data?

- Prostorové rozmístění obyvatelstva
 - Rozdíl mezi trvale a obvykle bydlícím obyvatelstvem
- Pohyb obyvatelstva
 - SLDB – vysoký podíl nevyplněných nebo částečně vyplněných odpovědí
 - SLDB – data za jeden „okamžik“
 - Další zdroje – sčítání v dopravě (MD)
- ...

Data z mobilních sítí

- Jak vlastně data z mobilních sítí fungují?
 - Data z volání
 - Data ze sms
 - Lokalizační data
- Znamená to, že nás může někdo sledovat?
 - Možnosti přesné lokalizace jednotky
 - Vysílač/BS – prostorové vymezení

Datová omezení

- Data pouze za „živé“ uživatele
- Žádná historie (unique hash user-id per 24h)
- Pouze 3 informace (user-id, time, cell-id)
- Žádná informace o hovoru/činnosti (handover, location update, etc.)

Type of records (2G)	
Location update	53%
Mobile terminating/ originating call/SMS	32%
Others	15%

Type of records (3G)	
Location update	62%
Mobile terminating /originating call/SMS	18%
Others	20%

Jak taková data vypadají?

Záznamy v databázi

d332afc6b33190d3ef00a1217996935a,2014-04-04,00:00:36,BPHLOU4,GBTDR

8c52a1e0f5179b3fc1d9a2e87699b303,2014-04-04,00:05:00,BD1HOR4,IUCS

8c52a1e0f5179b3fc1d9a2e87699b303,2014-04-04,00:04:59,BD1HOR4,GBTDR

...

Lokalizace vysílače

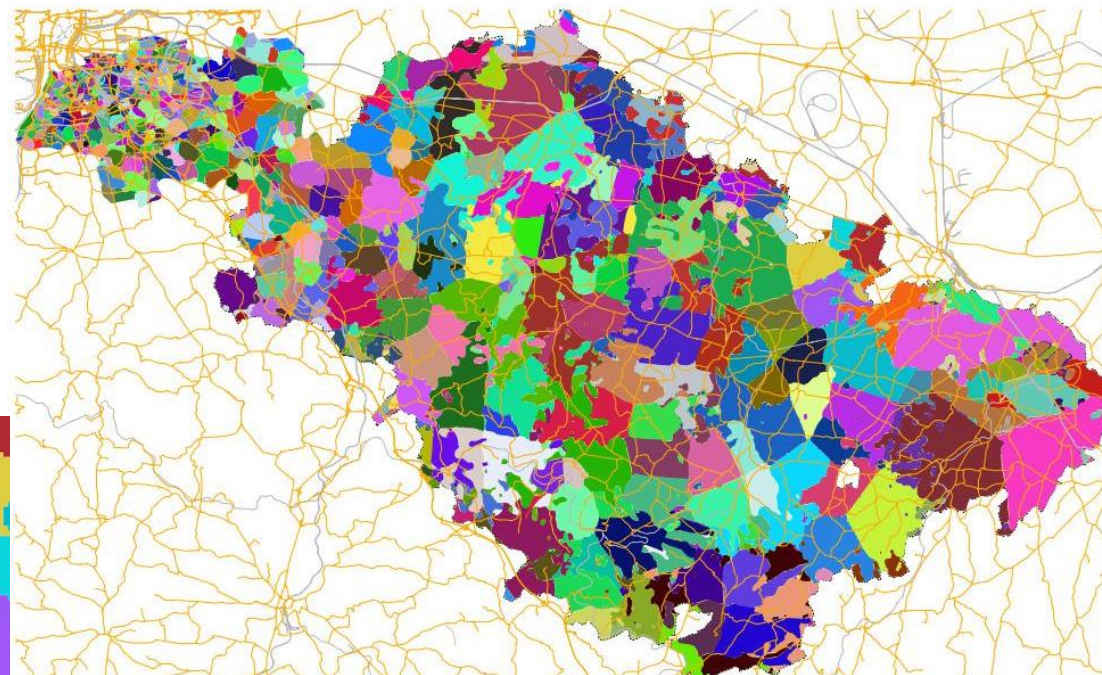
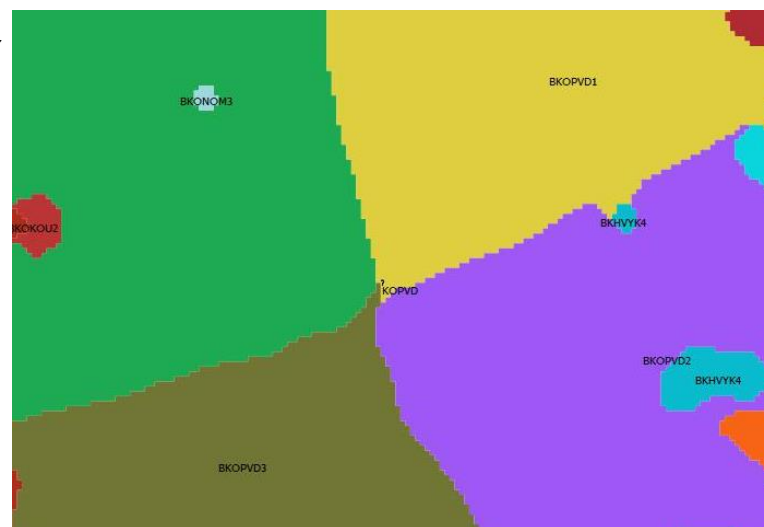
D1OST, 49.838375,14.827956

POZAH,50.054701,14.515403

PHSBO,49.939251,14.690412

...

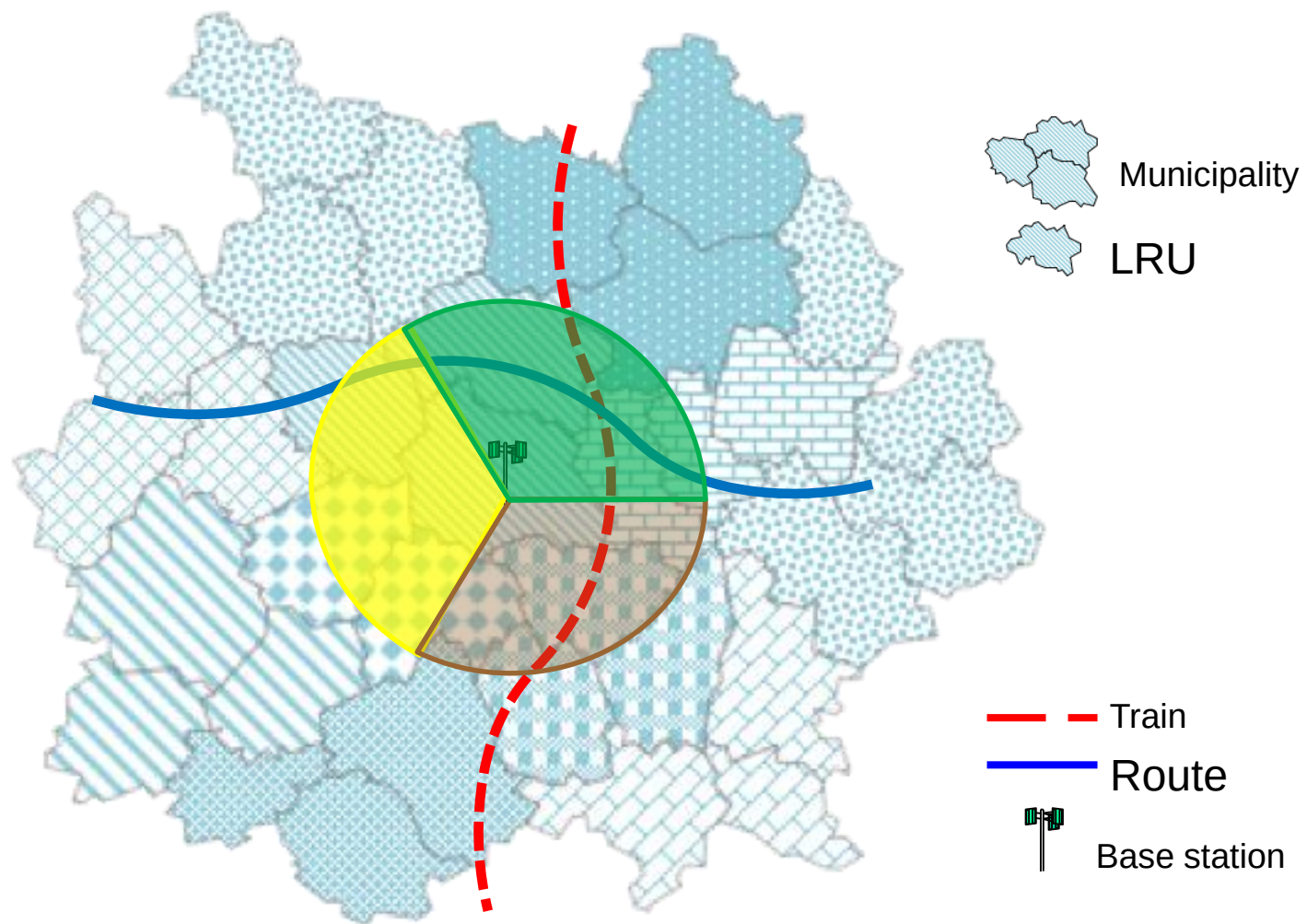
Seznam vysílačů a jejich pokrytí



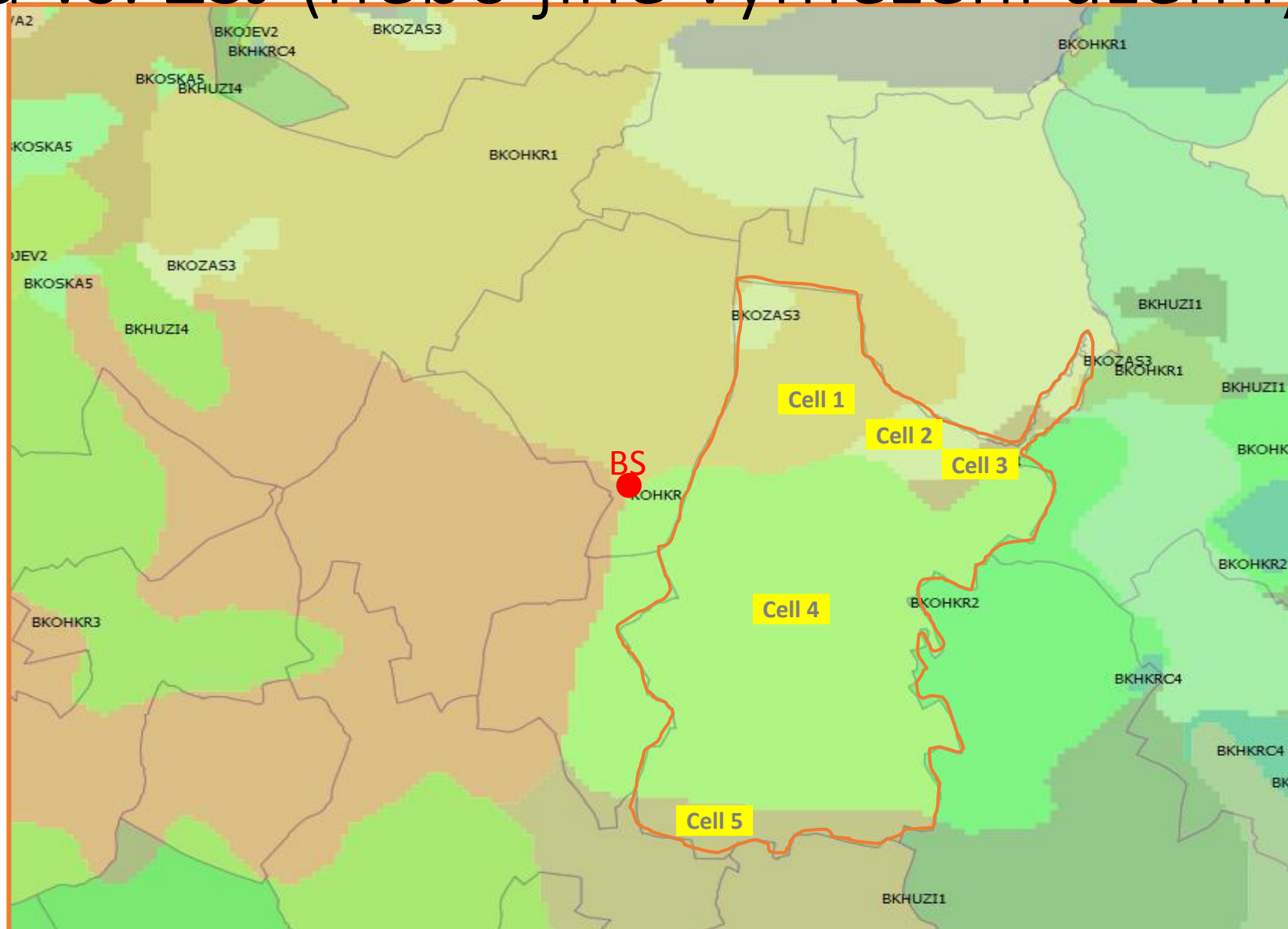
Jak velká data asi jsou?

	Per day
Počet záznamů	> 60 millions <i>(>1 Billion per 14 days)</i>
Buňka s nejvyšším počtem záznamů	34092
Buňka s nejnižším počtem záznamů	10
Počet buněk	1562
Počet detekovaných uživatelů	> 0.5 million

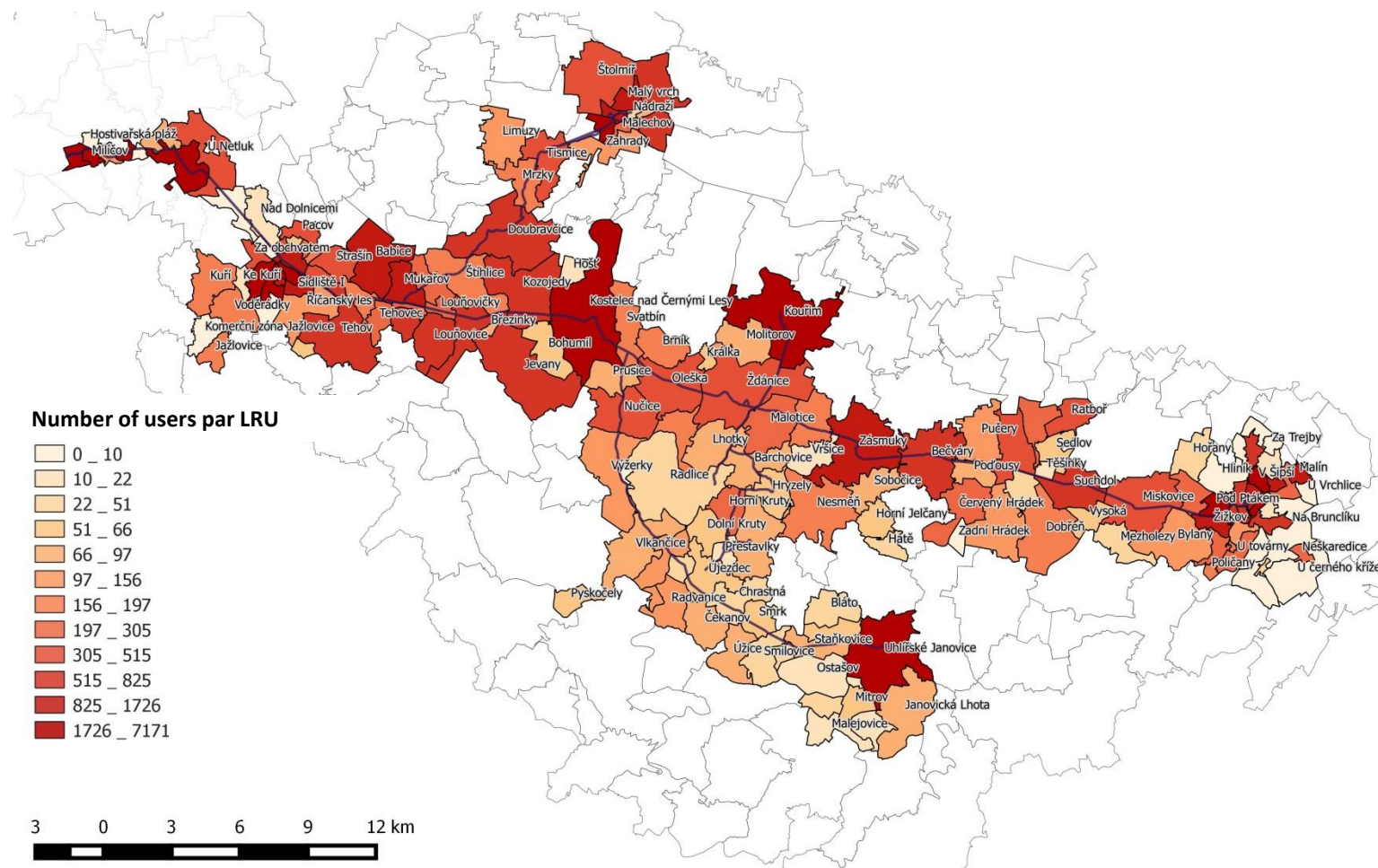
Buňka vs. ZSJ (nebo jiné vymezení území)



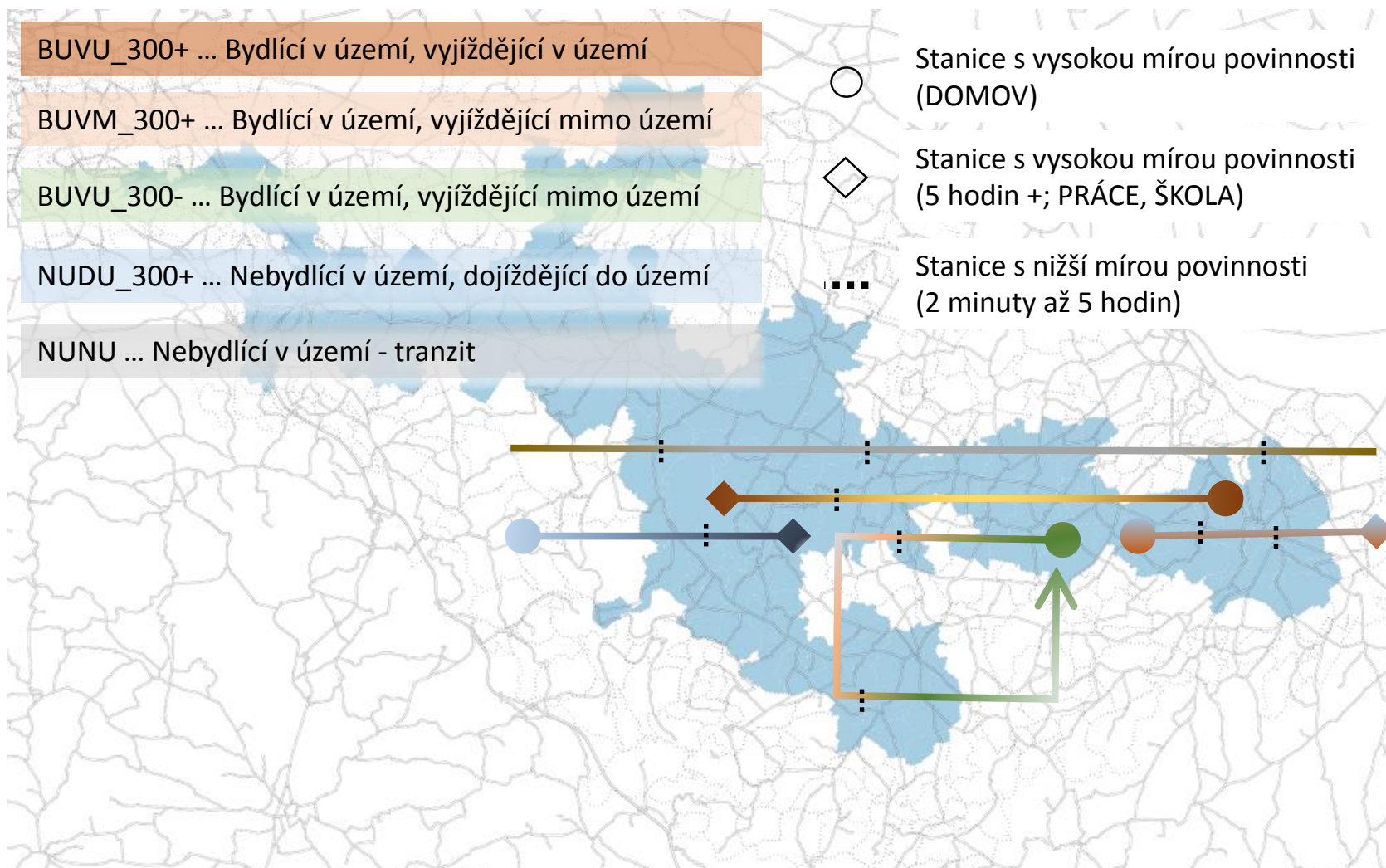
Buňka vs. ZSJ (nebo jiné vymezení území)



Počet uživatelů v ZSJ

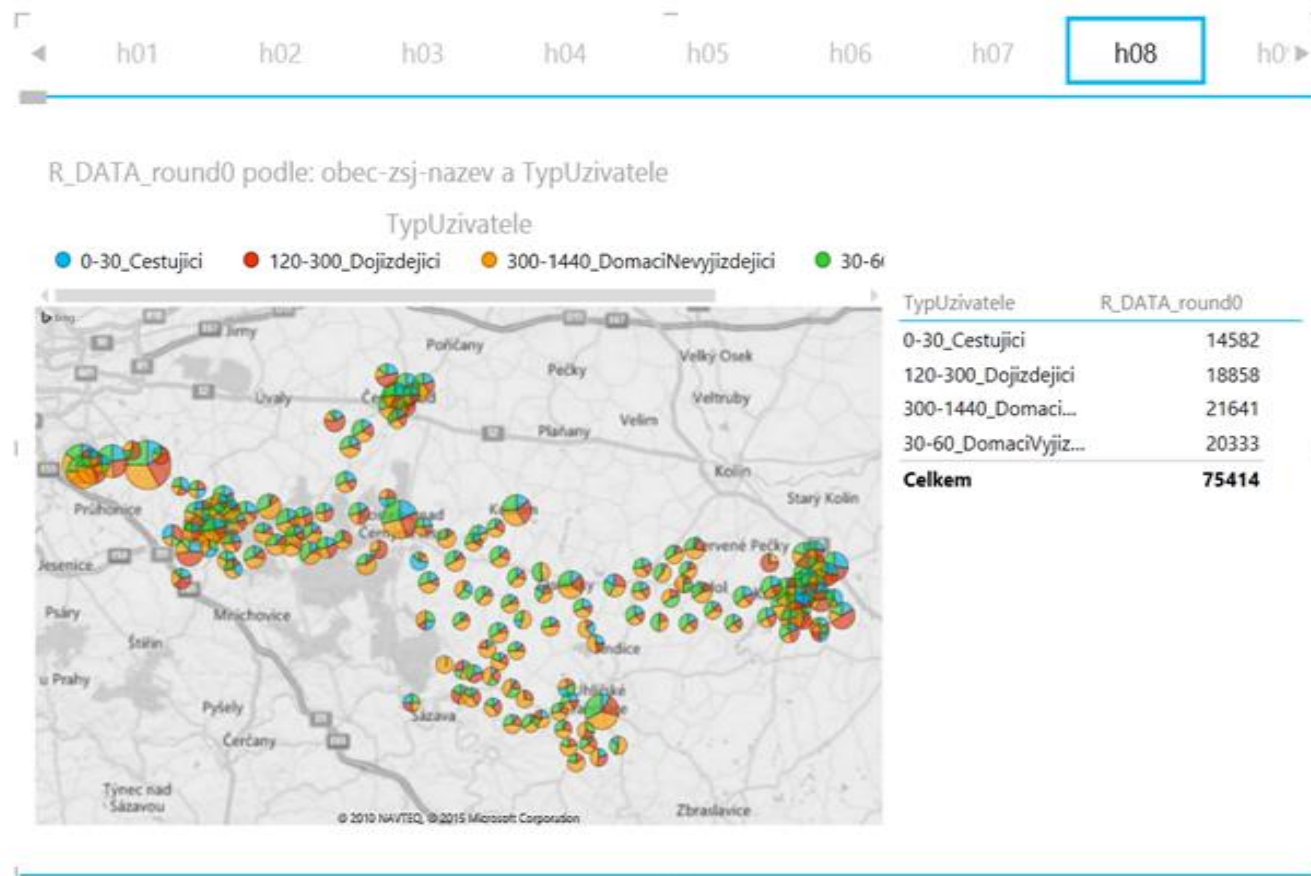


Počty osob ve stanicích - klasifikace

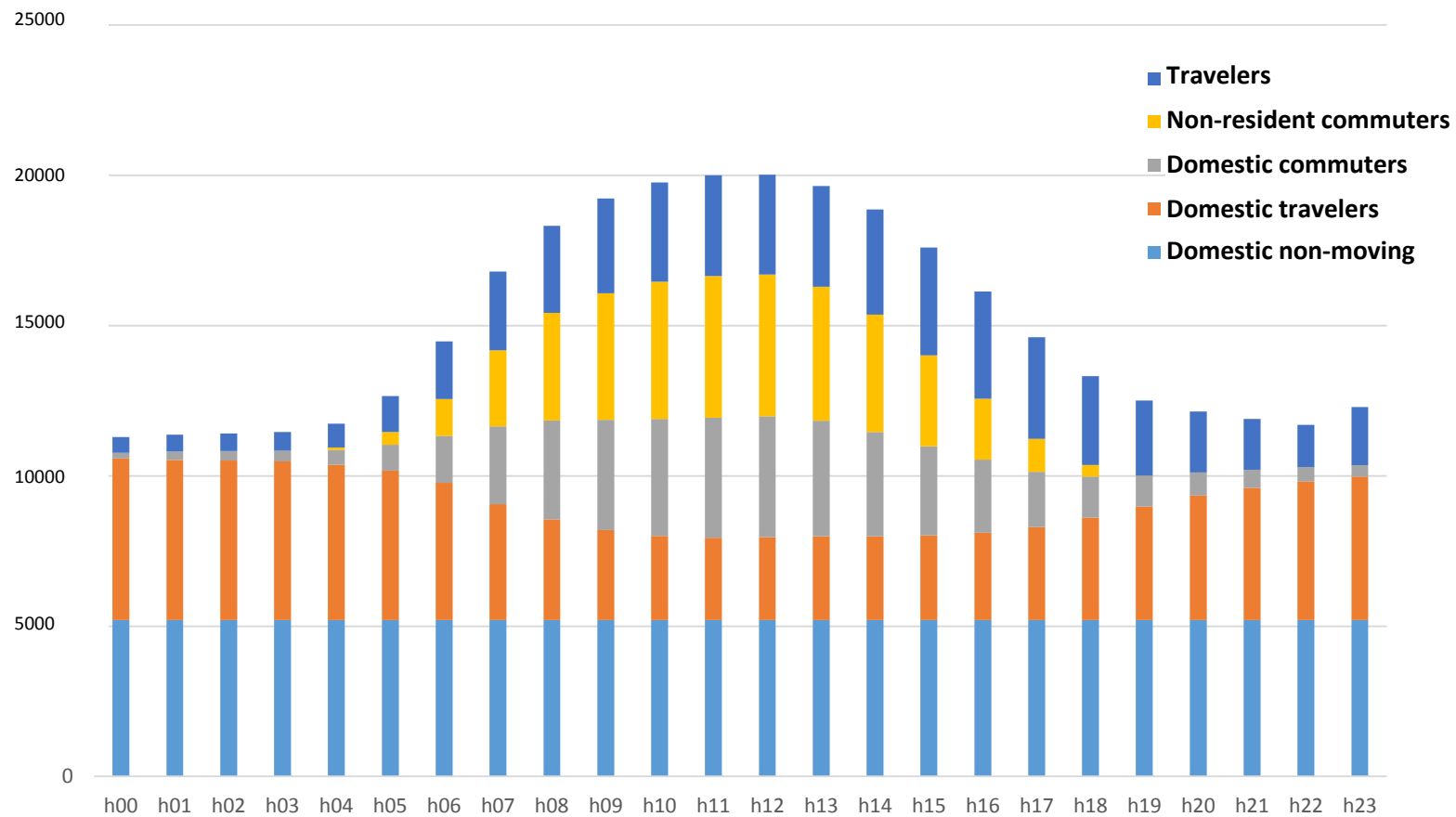


Počty osob ve stanicích dle typu

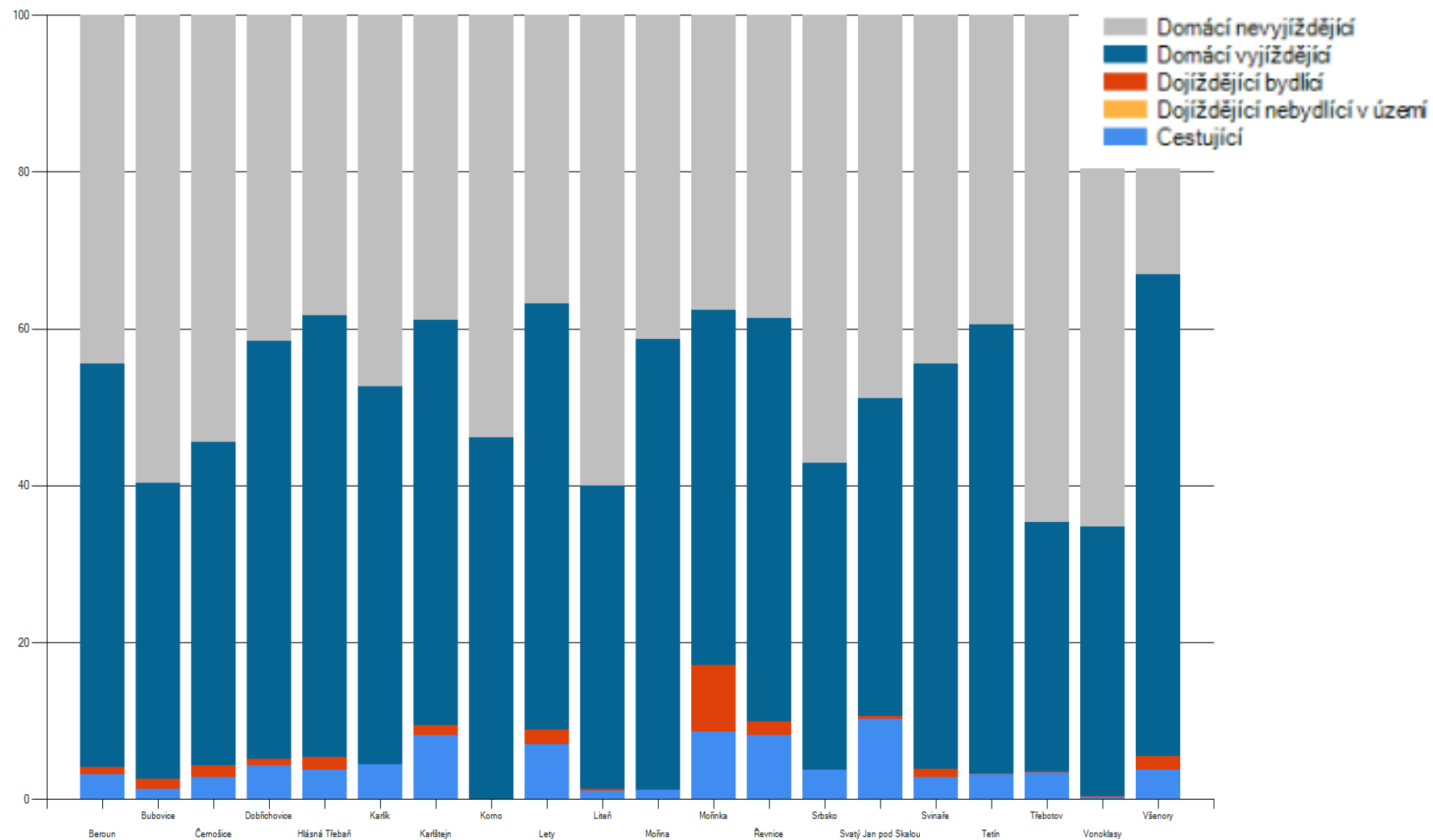
Počty uživatelů v zájmovém území dle typu a času výskytu



Klasifikace uživatelů

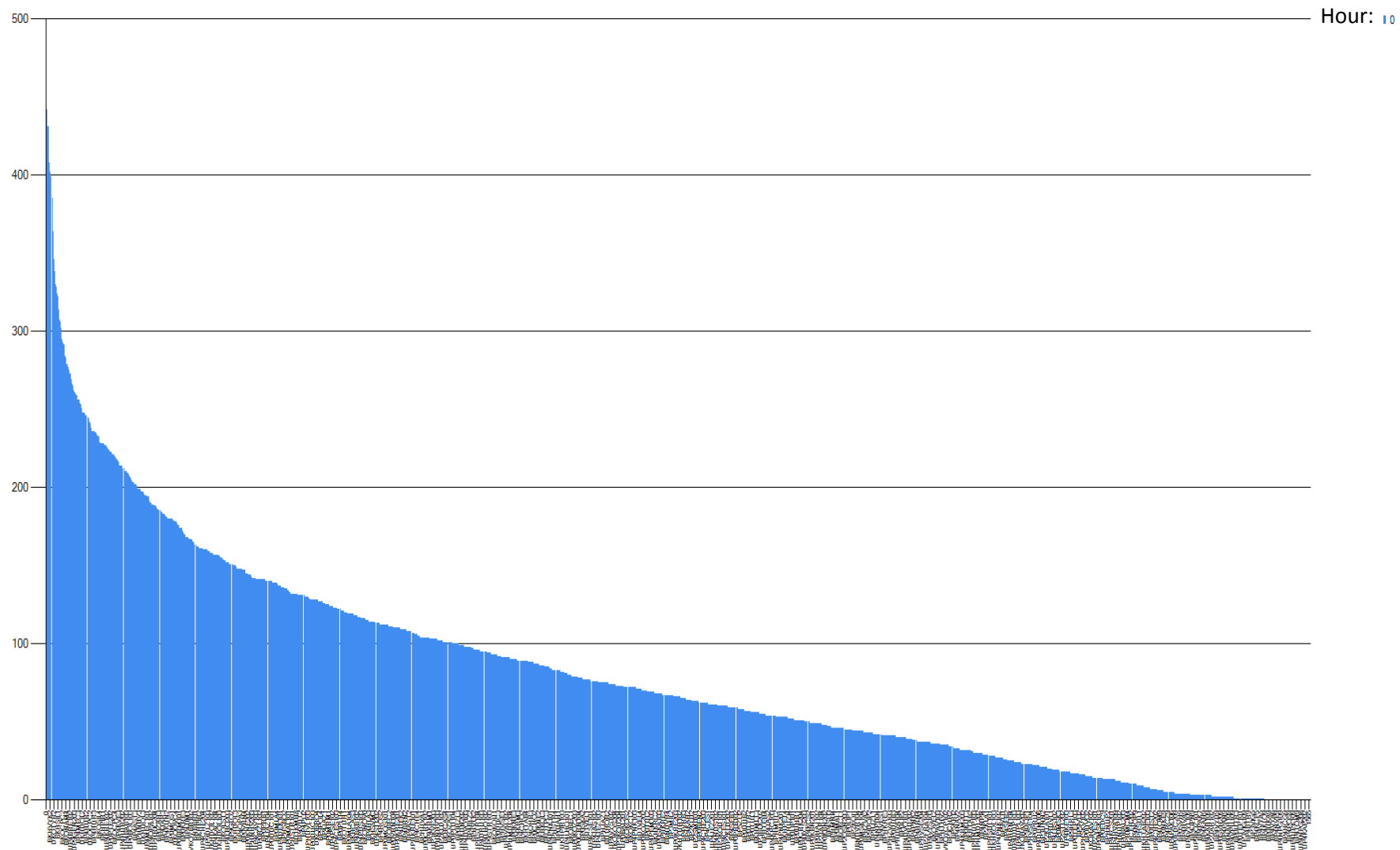


Vývoj počtu uživatelů podle typu v obcích

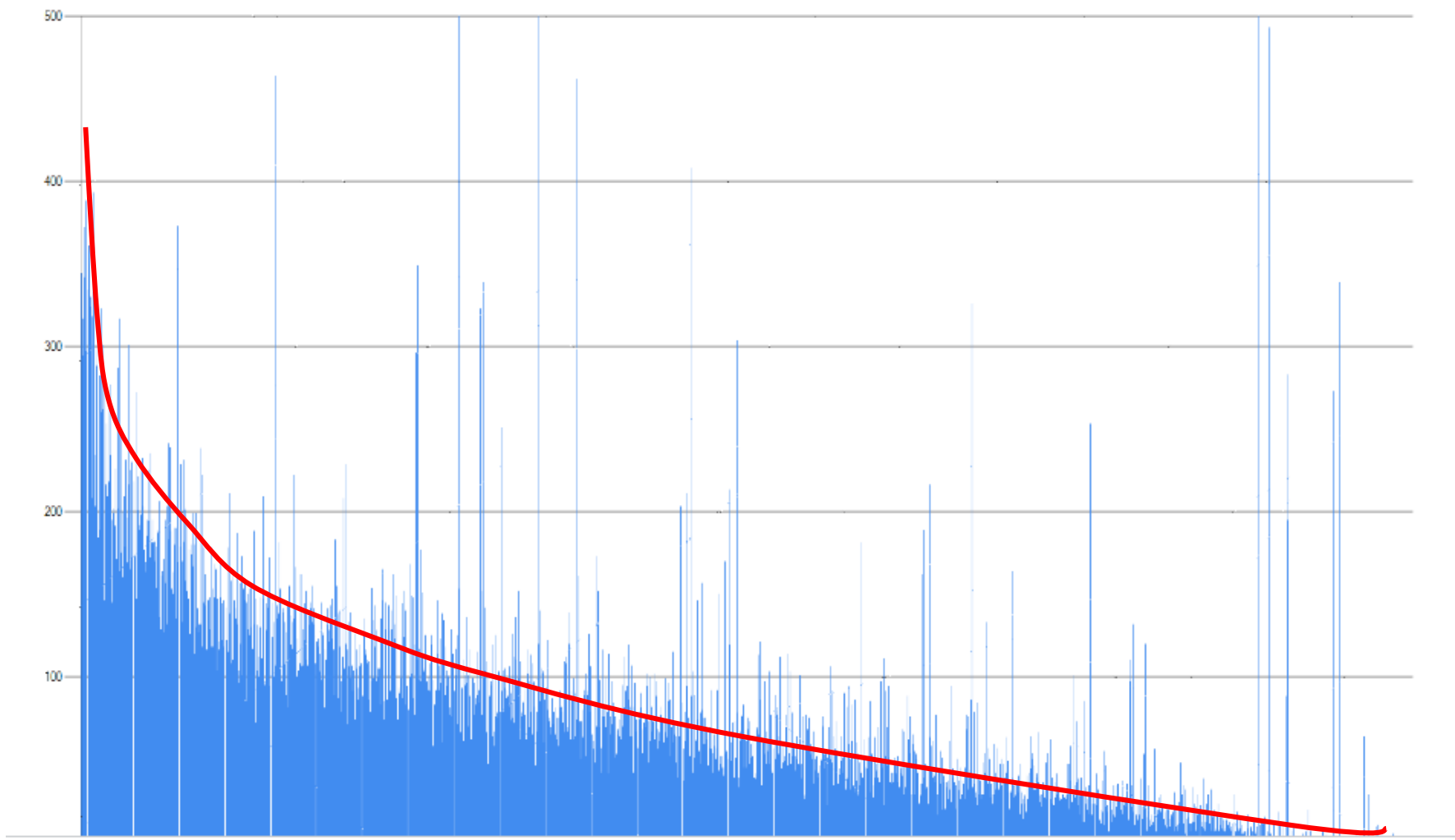


Vývoj záznamů v buňkách

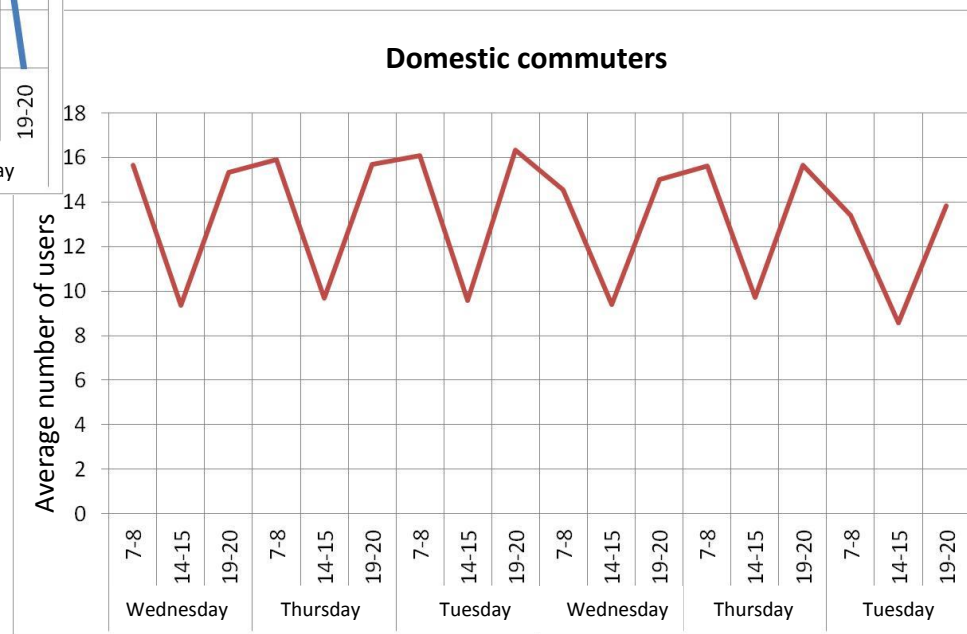
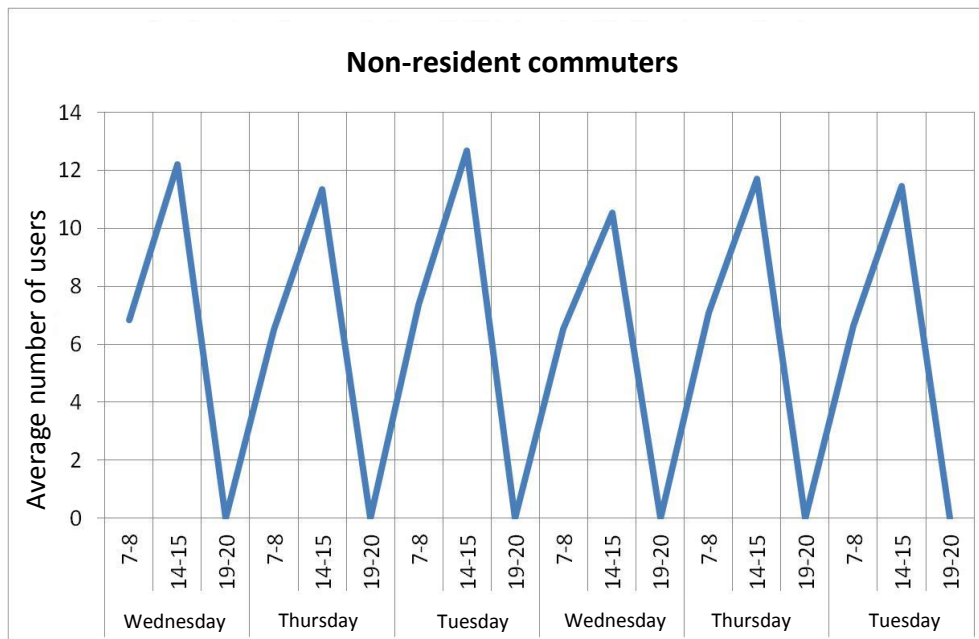
www.gif-animator.com - UNREGISTERED



Vývoji záznamů v buňkách – srovnání



Stabilita v čase



Výhody

1. Výskyty a přesuny obyvatel „tak jak jsou“ = „nová realita“
2. Územní detail, časový rozměr = uspokojení poptávky
3. Možnost zaměřit změny tam, kde je nejlepší poměr užitné hodnoty a nákladů (investice, intervence)
4. Rychlá a efektivní zpětná vazba, možnost korekce
5. Z pohledu přepravy, porozumění chování na úsecích

Současné a budoucí možnosti využití

- Omezené studie (počet zákazníků centra, počet turistů, apod.)
- Plánování dopravní sítě
- Plánování veřejné dopravy
- Zdroj pro odhad skutečně bydlícího obyvatelstva
- Možnost využití při odhadu sezonních výkyvů
- Odhad frekvence, intenzity a způsobu dopravy