

Pokročilé propočty

Agenda

- 1) Středočeské volby 2020
- 2) Místo "pro demonstraci"
- 3) Techniky prostorové ekonometrie
- 4) Optimalizace distribuční sítě
- 5) Diskuze

Dva základní koncepty

- Informace je v bodech (souřadnicích)
 - Typické pro přírodní veličiny / geodesie, minerály, počasí
 - Kriging / stochastický spojitý proces
 - Klíčem je vzdálenost
- Informace je v oblastech (polygonech)
 - Typické pro ekonomické veličiny / admin areas
 - Spojitost jevu nedává smysl; identifikujeme (auto)korelaci
 - Klíčem je sousedství

Pár slov k AI



- Neuronové sítě = pattern recognition
- Vysoký nárok na objem trénovacích dat – řádově desetitisíce pozorování
- Existují třídy problémů, kde objem dat dává smysl – object recognition (počítání lidí / aut na fotce)
- Pro ekonomická (ekonometrická) data spíše méně pravděpodobné

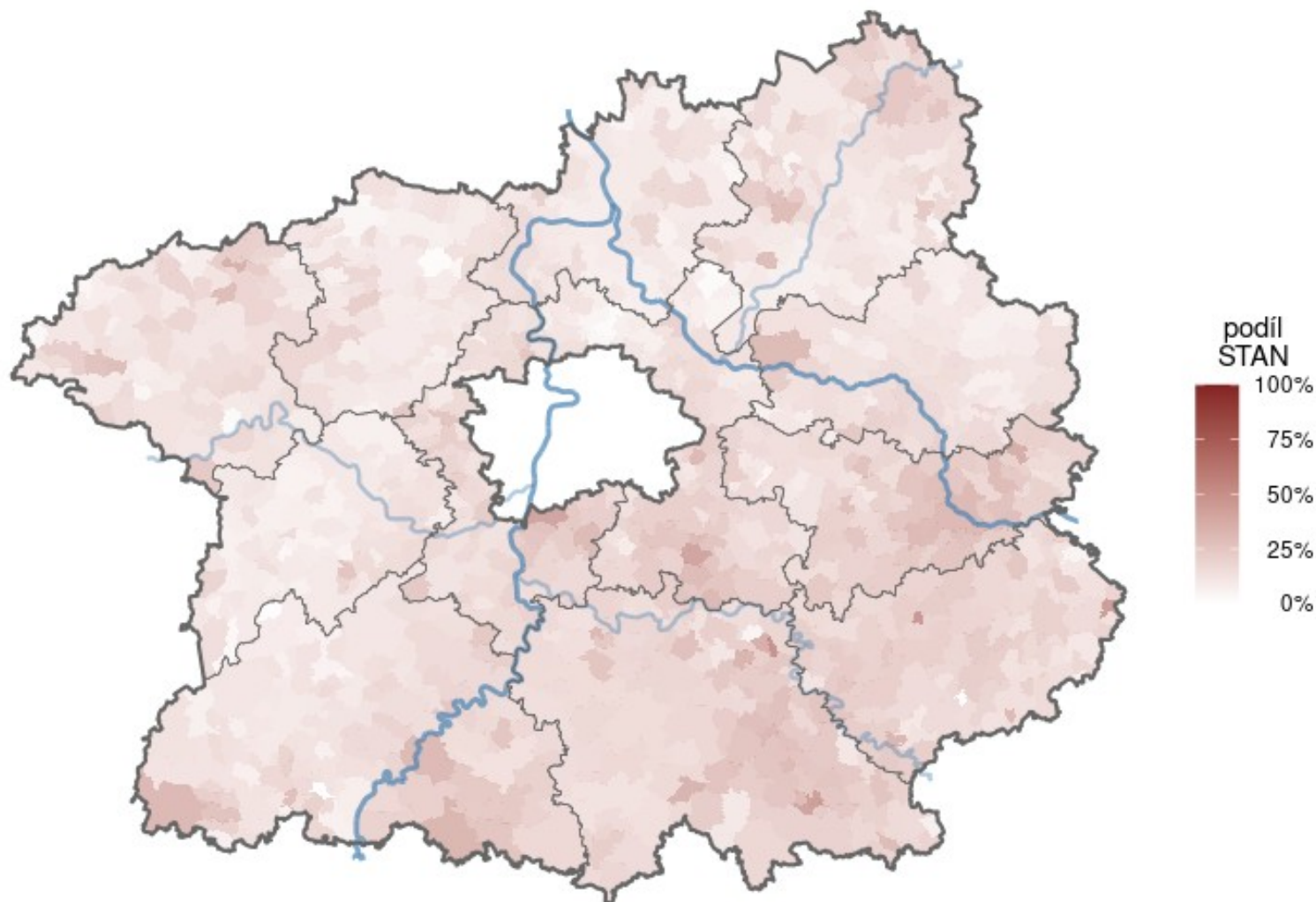
Středočeské volby

- Zdroje dat:
 - Výsledky (csv) z ČSÚ
 - Polygony volebních okrsků z ČÚZK
- Spojení dat / filtrace 1 kraje
- Základní přehled

Relativní hodnoty

Středočeské volby 2020
obecní výsledky / relativně

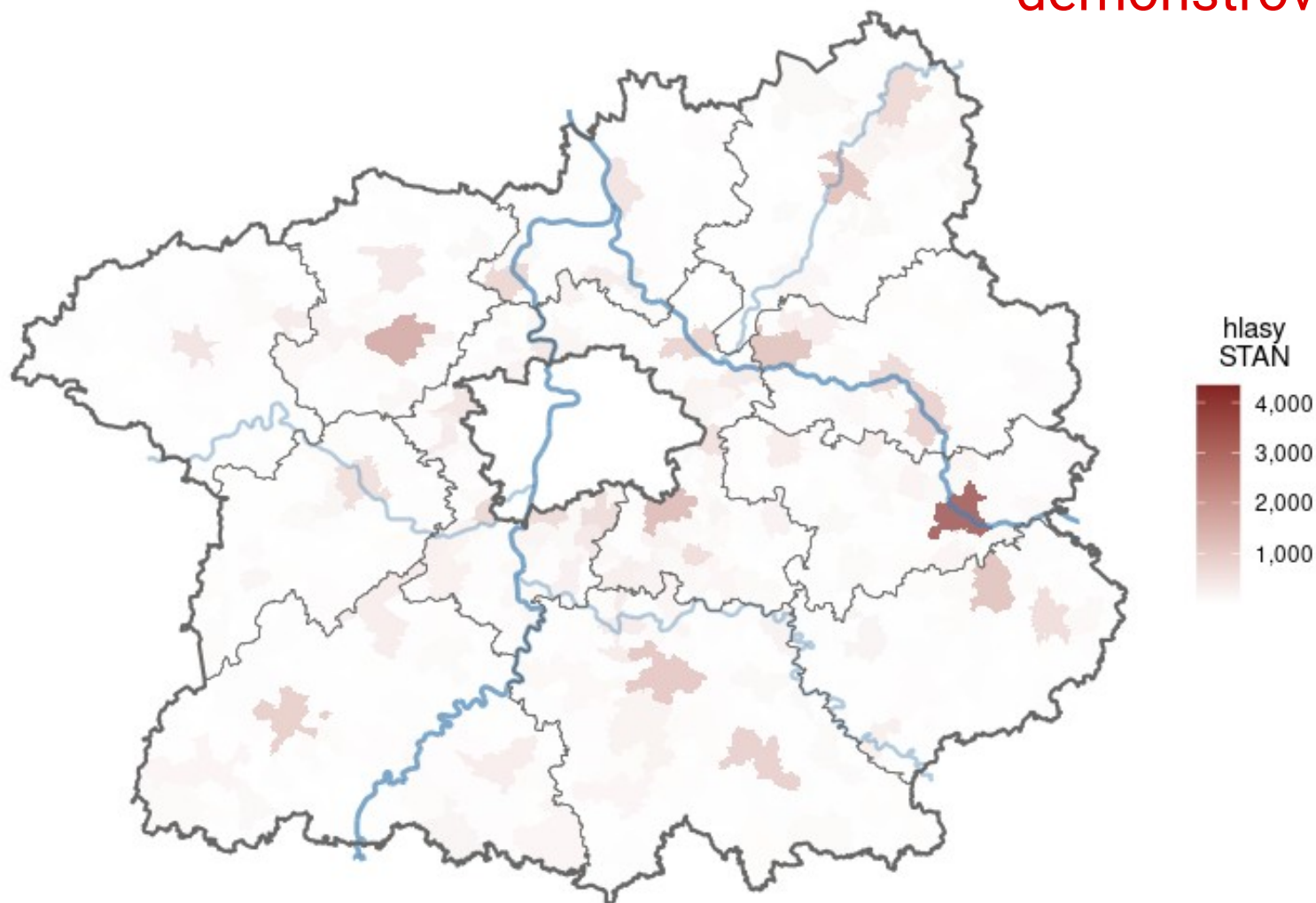
Tohle bývá v novinách...



Absolutní hodnoty

Středočeské volby 2020
obecní výsledky / absolutně

Lidé, ne procenta, chodí
demonstrovat!



Řešený příklad

4-digest-data-STČ.R
4-okrsky-viz.R

Problém "demonstrace"

- Cíl optimalizace: najít bod, v jehož okolí je nejvíce příznivců (voličů)
- Řešení hrubou silou:
 - Buffer o okruhu n metrů
 - Kolem všech bodů / součet
 - Vybrání maxima z úplné množiny

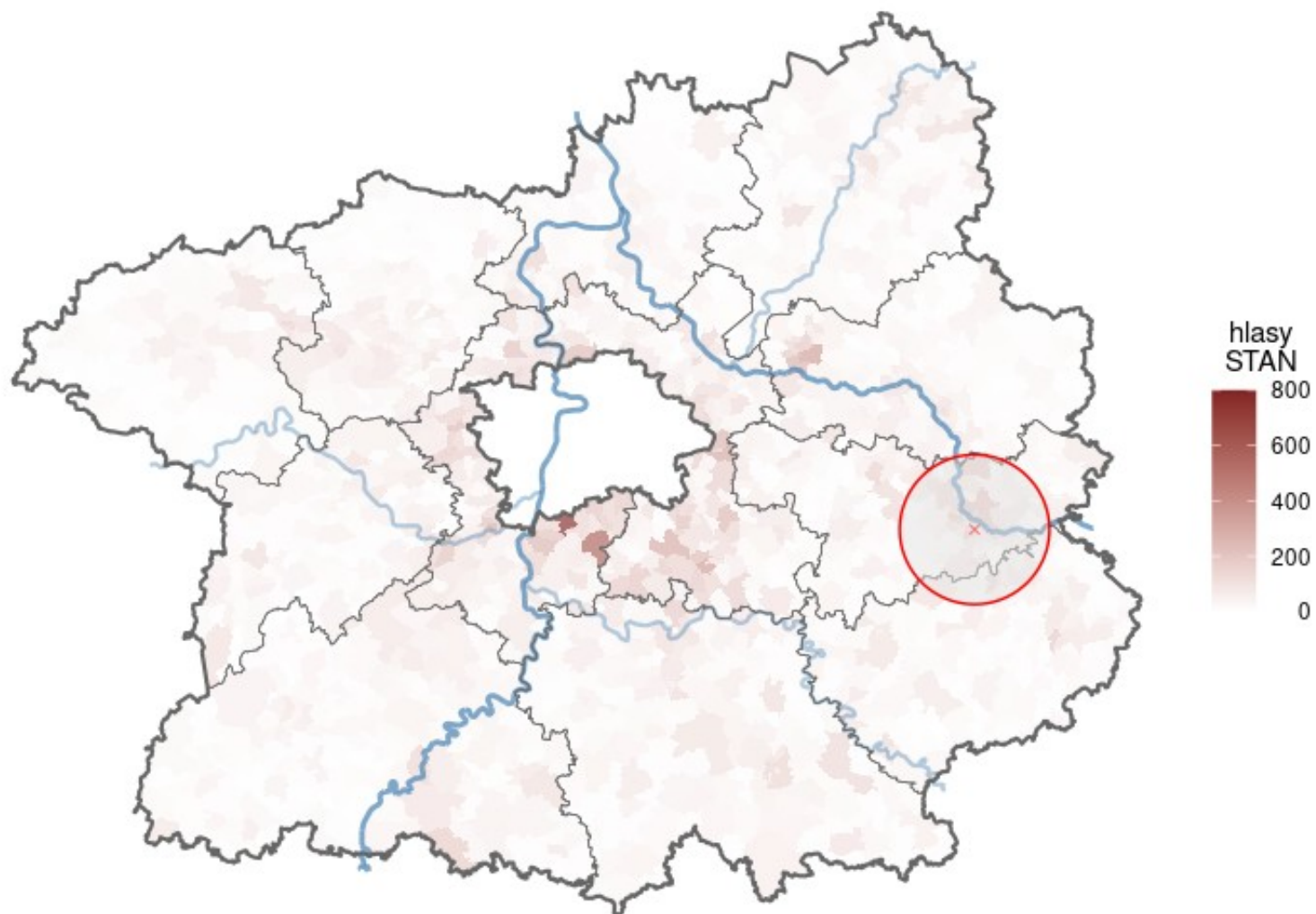
Řešený příklad

4-demo-hrubou-silou.R

Demo 10 km

Optimální sraz pro voliče z okolí 10 kilometrů

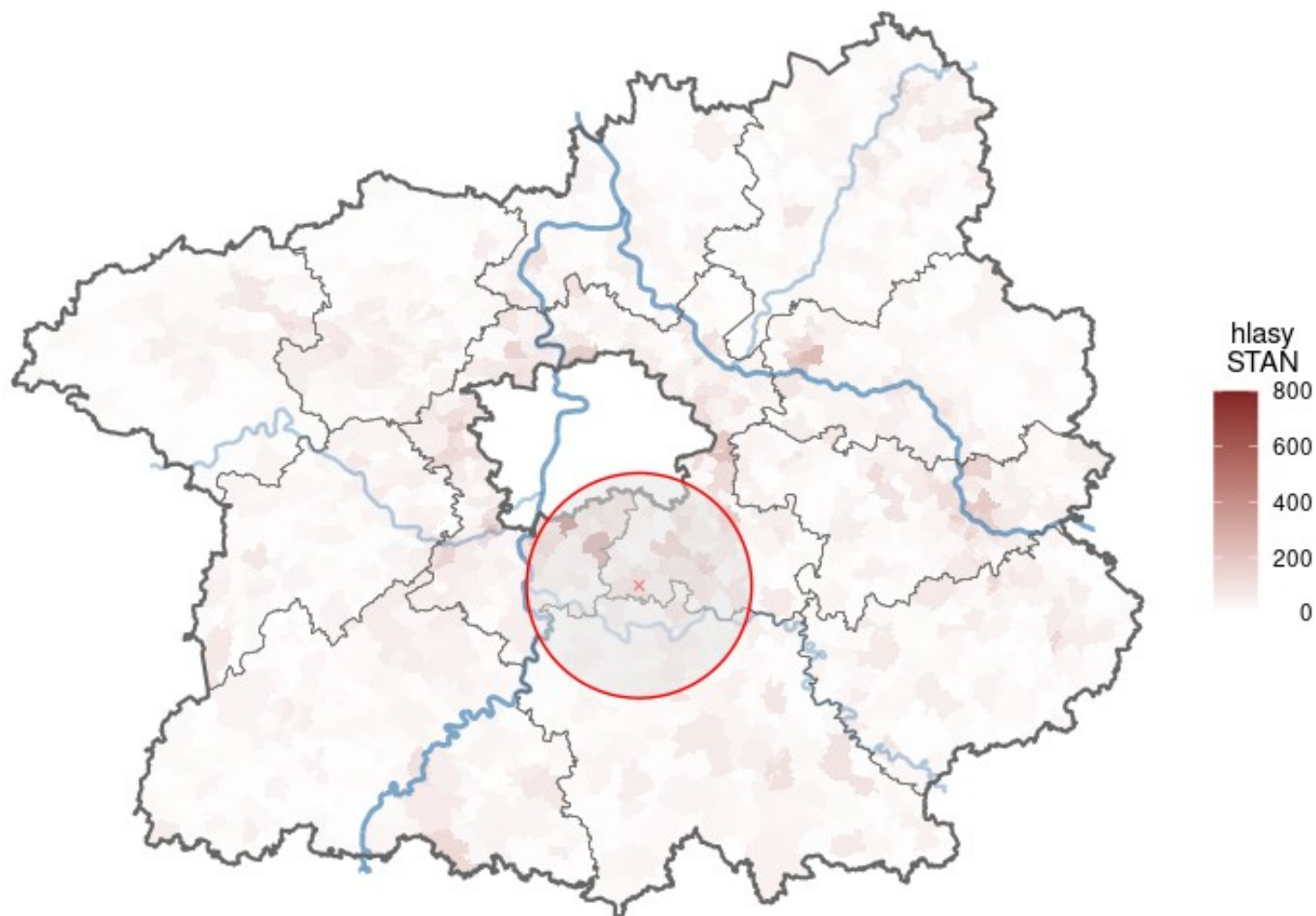
Orebitská 885, Kolín IV, 28002 Kolín



Demo 15 km

Optimální sraz pro voliče z okolí 15 kilometrů

Ládví č.ev. 1473, 25168 Kamenice



Prostorová ekonometrie

- Hypotéza a test
- Je veličina náhodně rozmístěná?
- Co znamená náhodně rozmístěná?
= je prostorově (auto)korelovaná?
- Pokud není: clustery (oblasti se sobě podobným chováním)

Přístupy ke clusterům

- `{spdep}` od Rogera Bivanda (Bergen)
- `{rgeoda}` interface k funkcím GEODA od Luca Anselina (Chicago)

Přístupy ke clusterům

- Test rovnoměrného rozložení (autokorelace) Moranovo I

```
spdep::moran.test()
```

- Lokální Moran I hodnoty

```
spdep::localmoran()
```

- Lokální Getis Ord G^* hodnoty

```
spdep::localG()
```

Interpretace clusterů

- Výhoda: jsou neoddiskutovatelné
- Nevýhoda: korelace není kauzalita
- Problém: přenos na jiné regiony
- Zpravidla maskují v datech skrytou veličinu
- U STAN je to starostování v obcích

Řešený příklad

4-spdep.R
4-geoda.R

Optimalizace

- Optimalizace distribuce
- Hledání závislosti (regrese)
- Zvláštní pozornost – rezidua / rozdíl mezi modelem a skutečností
- Regrese = standardní statistické techniky – `stats::lm()` a přátelé

Optimalizace

- první krok: společný jmenovatel
 - dobrý start: mřížka (odstíní rozdíly v ploše polygonů)
 - administrativní polygony problematické
- druhý krok: klasická statistika
 - může, a nemusí, být prostorový lag

Optimalizace

- Modelový příklad: optimalizace hospod v Praze
- Faktory
 - Počet obyvatel městské části
 - Počet hotelových lůžek
 - Kvalita vegetace (zeleň)

Řešený příklad

4-optimalizace-*.R

Diskuze...