

Pokročilé počty

$$\frac{n \times \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \times \sum_{i=1}^n (x_i - \bar{x})^2}$$

Agenda

- 1) Prezidentské volby 2023
- 2) Techniky prostorové ekonometrie
- 3) Optimalizace distribuční sítě
- 4) 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
 - Pracujeme s agregáty = spojitost jevu nedává smysl
 - Identifikujeme (auto)korelaci
 - Klíčem je sousedství (cokoliv to znamená)

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, segmentace...
- Pro ekonomická (ekonometrická) data spíše méně pravděpodobné

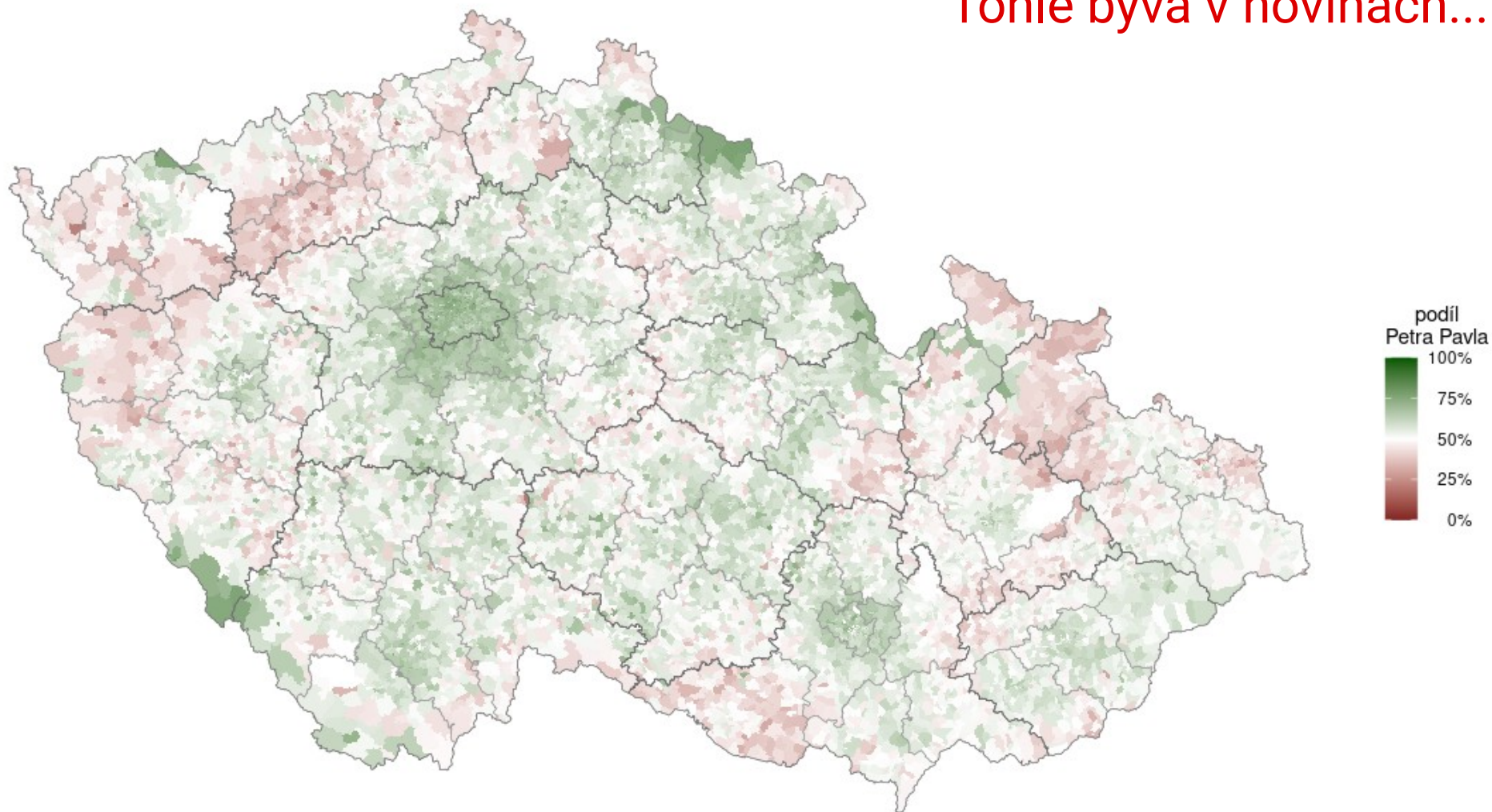
Prezidentské volby

- Zdroje dat:
 - Výsledky (csv) z ČSÚ
 - Polygony volebních okrsků z ČÚZK
- Doplnění o okresy & kraje
- Základní přehled

Relativní hodnoty

Prezidentské volby 2023

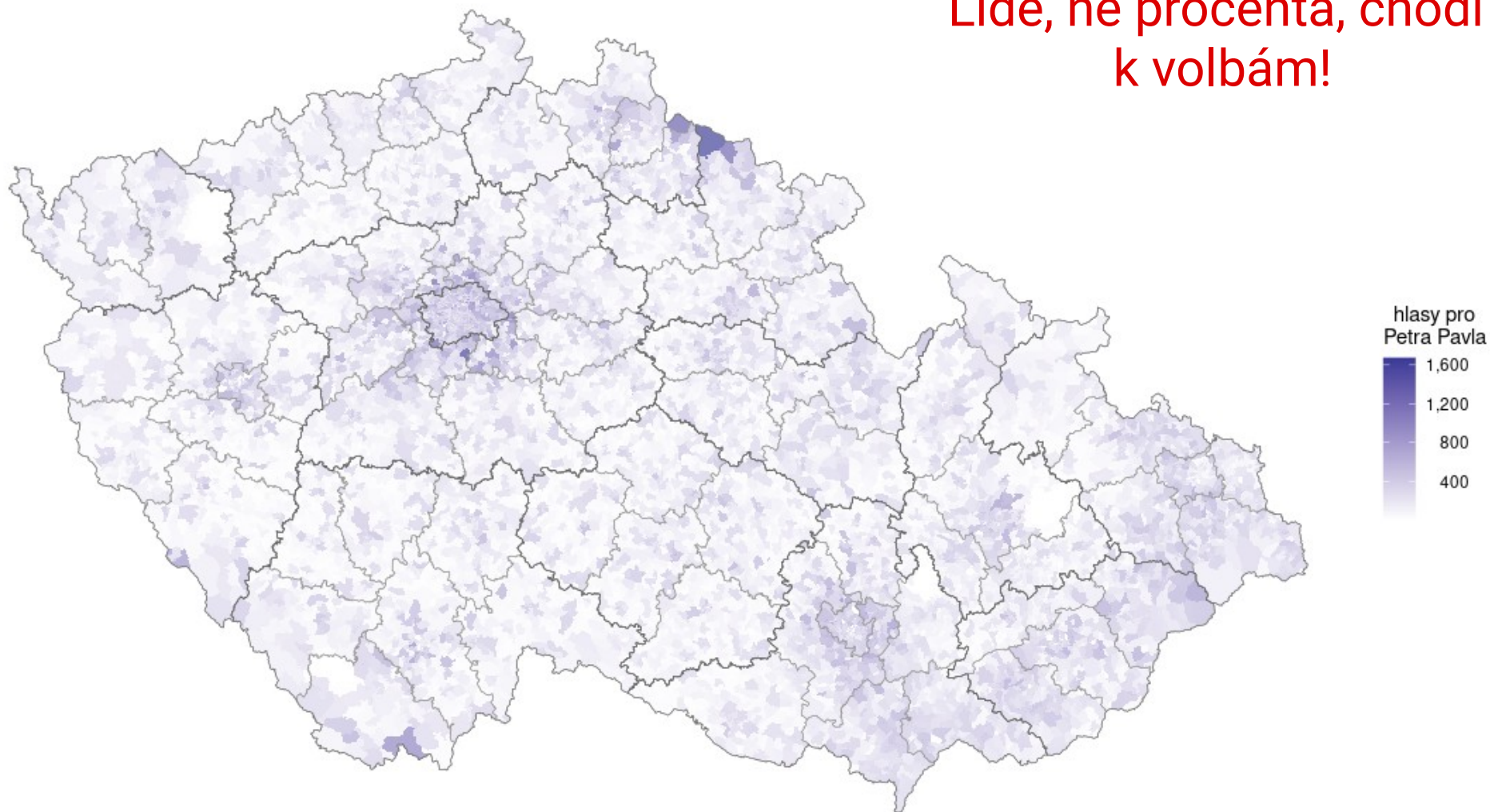
Tohle bývá v novinách...



Absolutní hodnoty

Prezidentské volby 2023

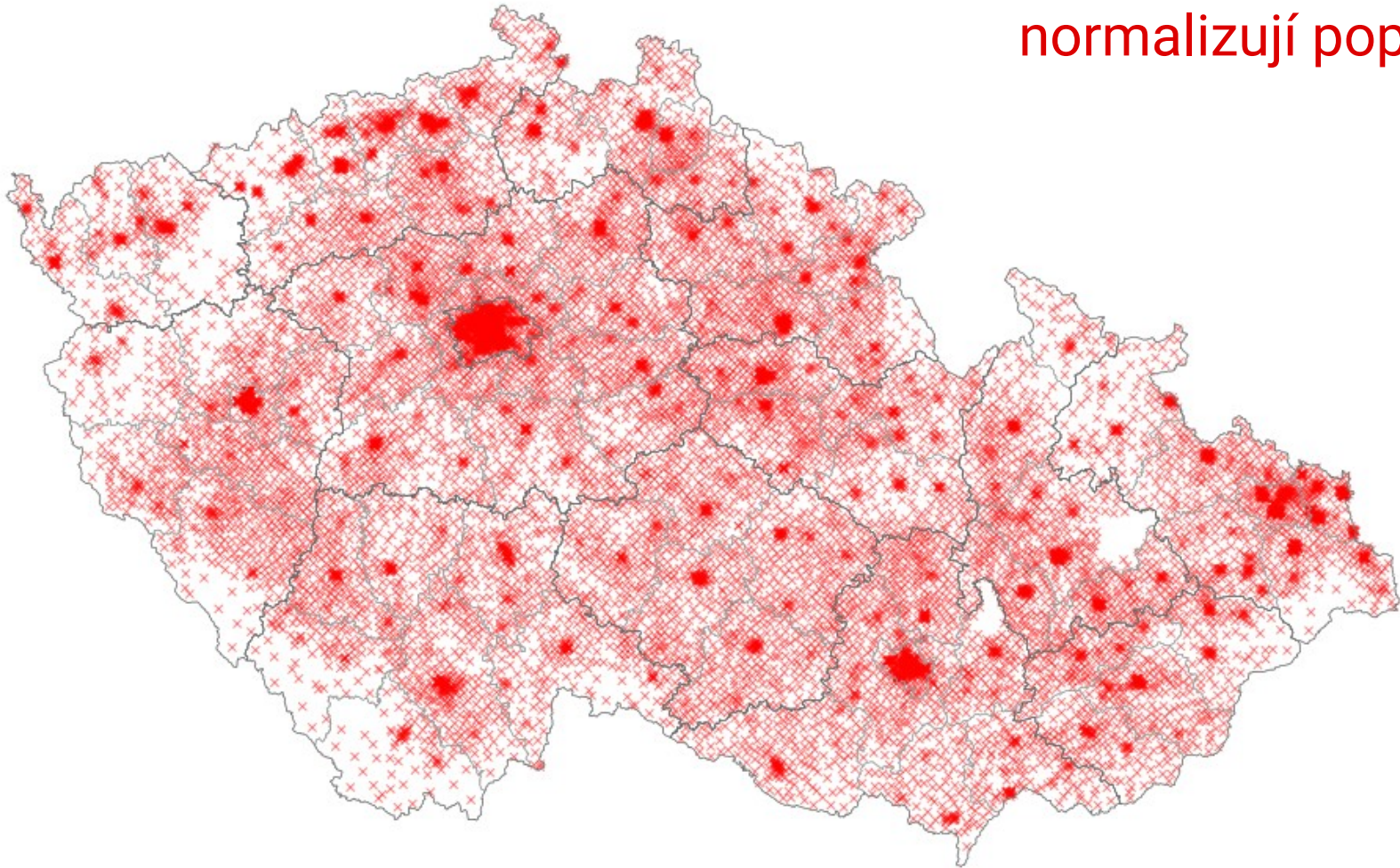
Lidé, ne procenta, chodí
k volbám!



Obvody a populace

Prezidentské volby 2023
středové body okrsků

Srovnatelně velké okrsky
normalizují populaci



Řešený příklad

4-digest-data-prez.R
4-okrsky-prez.R

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()`

Základní koncept Morana

$$I = \frac{n \times \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \times \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Uplatňujeme váhy na korelační koeficient
- Váhy nesou prostorovou interakci
- Hodnota / jde standardizovat na z skóre - které má normální rozdělení (s dopady)

Varianty vah

- **Sousedství:** čárka za společnou hranici, jinak nic
- **Vzdálenost:** zpravidla převrácená hodnota (převrácená hodnota vzdálenosti² / model gravitace)
- **Interakce:** vlastní analýza – obchodní výměna, délka společné hranic

Jak určit sousedství?

- Šachová **královna**: 1 bod společné hranice
- Šachová **věž**: společná úsečka, bod nestačí
- **KNN**: nejbližších n objektů, kde n je parametr (garance právě n sousedů)
- **Vzdálenost**: vše v okolí m , kde m je parametr
- Ve světě `{spdep}` KNN a vzdálenost očekávají body, věž a královna polygony

Šach pro nešachisty



sousedství
věže

FALSE
TRUE

Realizace v {spdep}

- `nb` objekt: kdo sousedí s kým (jako index objektů / sparse matrix)
- `listw` objekt – více složek [2]:
 - kdo sousedí s kým (`neighbours`)
 - jako vahou (`weights`)
 - je zvykem, aby řádky vah sčítaly do 1 (tj. \sum matice vah = počet řádků)

Řešený příklad

4-weight-matrices.R

Interpretace clusterů

- Výhoda: jsou neoddiskutovatelné
- Nevýhoda: korelace není kauzalita
- Problém: přenos na jiné regiony
- Zpravidla maskují v datech skrytou veličinu

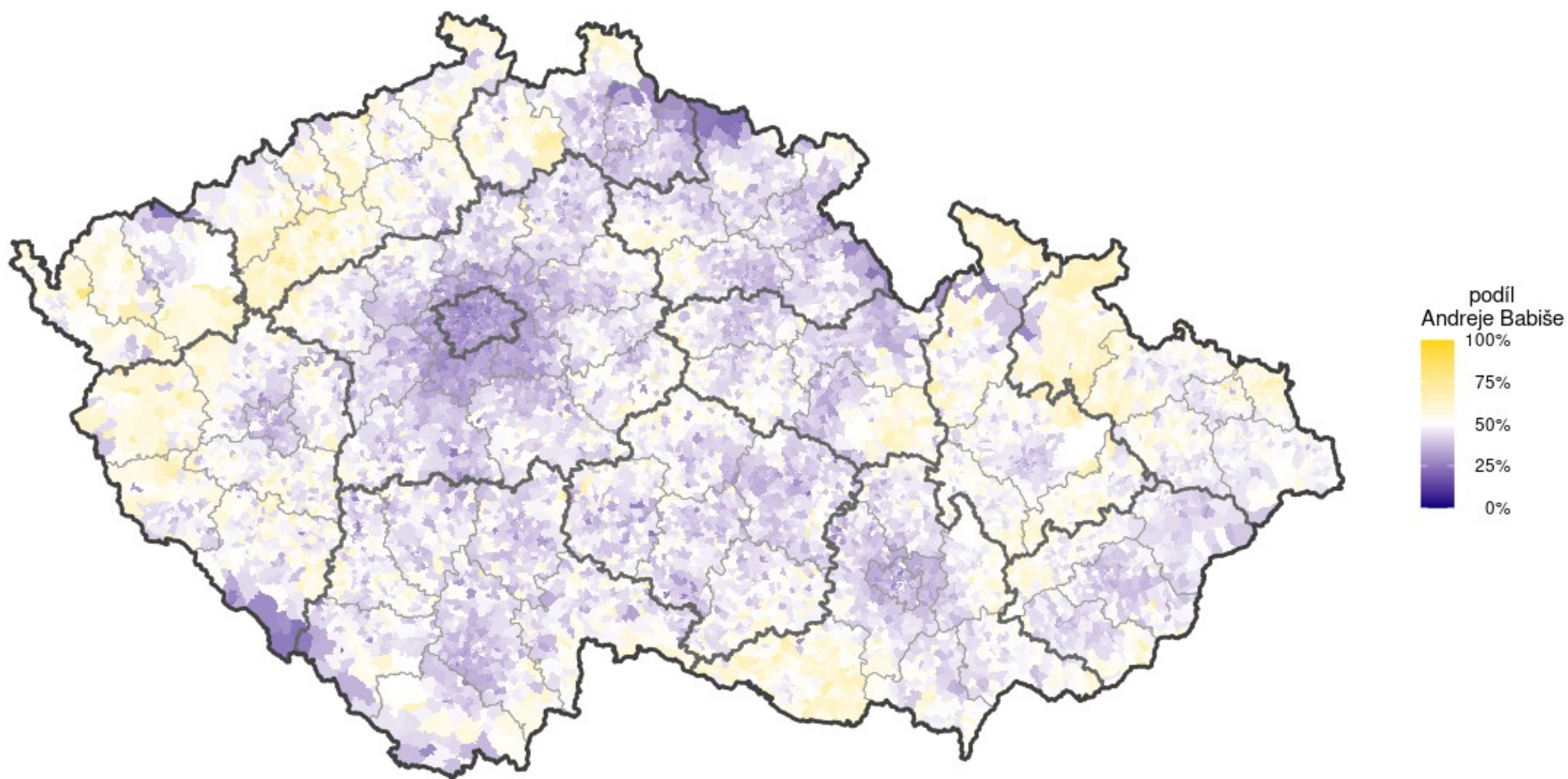
Řešený příklad

4-spdep-prez.R

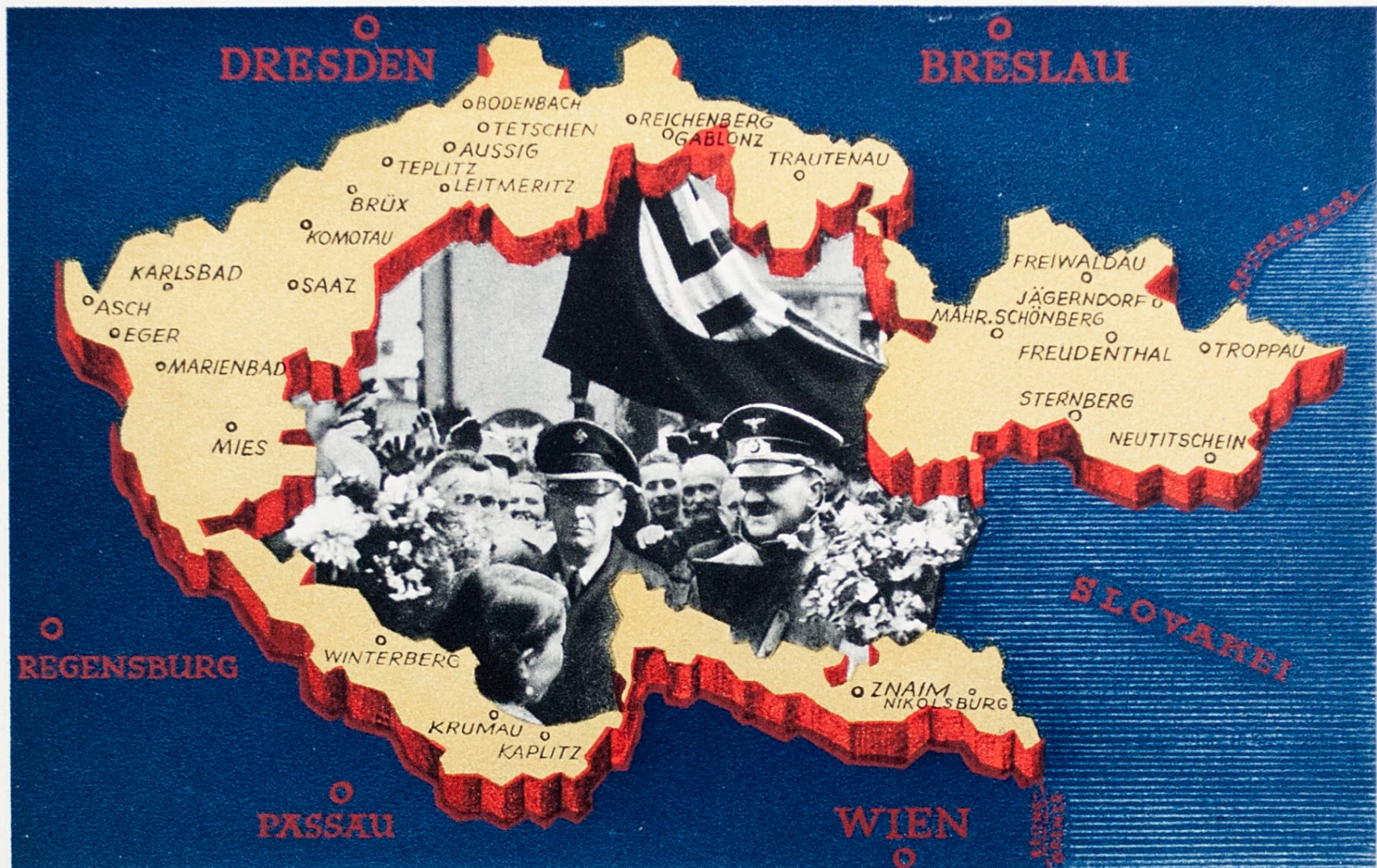
Příklad skryté veličiny

Prezidentské volby 2023

Připomíná vám tohle něc



Příklad skryté veličiny



WIR DANKEN UNSERM FÜHRER

Babiš a Sudety

- Zdroje dat:
 - Výsledky (csv) z ČSÚ
 - Sčítání lidu z roku 1930
- Přenos hlasů z roku 2023 na soudní okresy roku 1930 [n = 330]
- Klasická statistika...

Řešený příklad

4-regrese-prez.R

Optimalizace

- Optimalizace distribuce
- Hledání závislosti (regrese)
- Zvláštní pozornost – rezidua / rozdíl mezi modelem a skutečností
- Regrese = standardní erkové 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...