

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) Techniky prostorové ekonometrie
- 2) Prezidentské volby 2023
- 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

Pár slov k AI (ne LLM)



- 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

Koncept vah

- Matice sousedství **C** = kdo sousedí s kým? Dává výčet prvků.
- Matice vah **W** = jakou měrou sousedí kdo s kým? Vlastní váha.
- Matice hraje analogickou roli jako perioda v časových řadách

Varianty vah

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

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

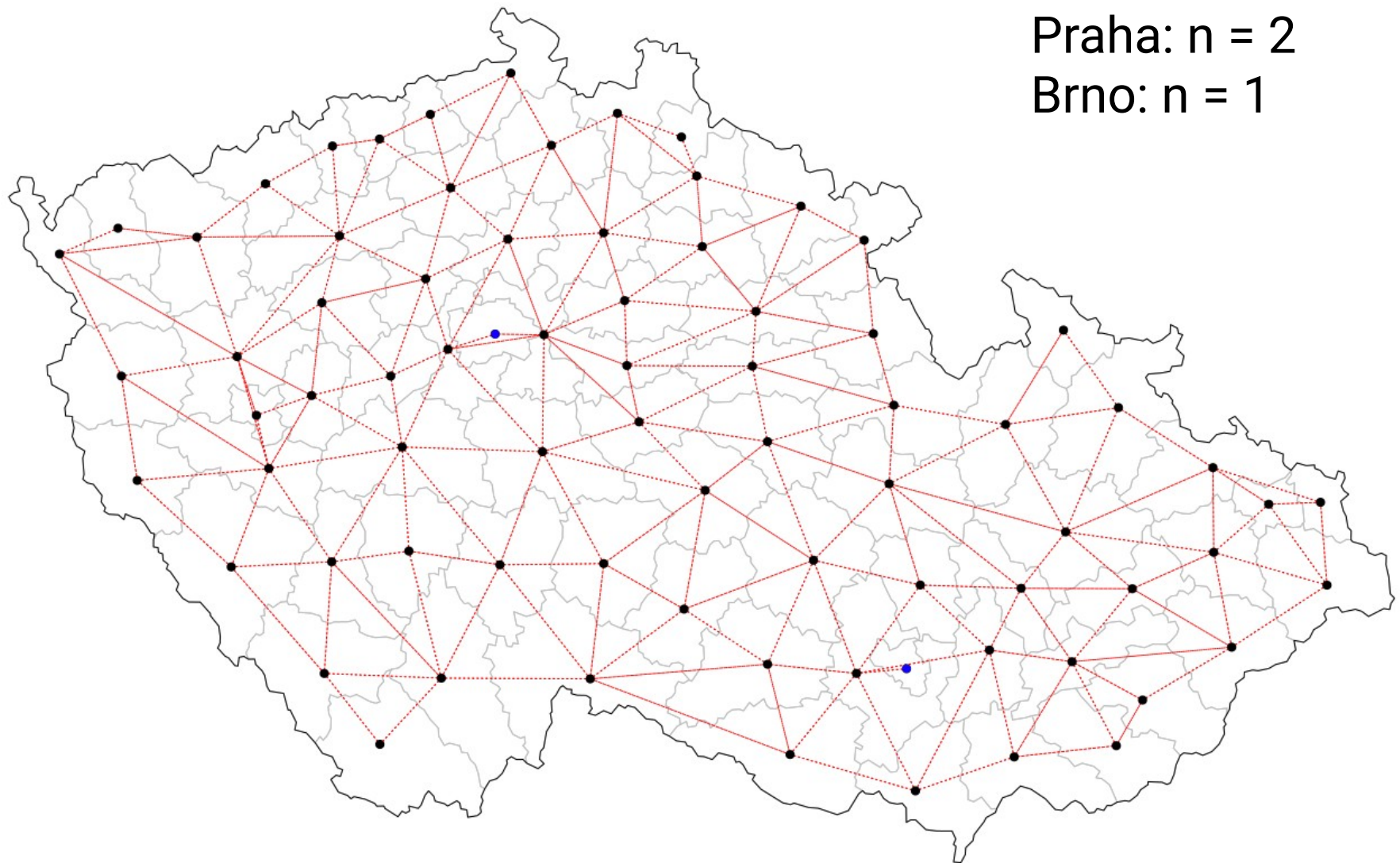
Realizace v {spdep}

- `nb` objekt: kdo sousedí s kým (jako index objektů / sparse matrix)
- `listw` objekt – více složek [3]:
 - kdo sousedí s kým (`neighbours`)
 - jako vahou (`weights`)
 - styl váhové matice – W je dobrý default (tj. Σ matice vah = počet řádků)

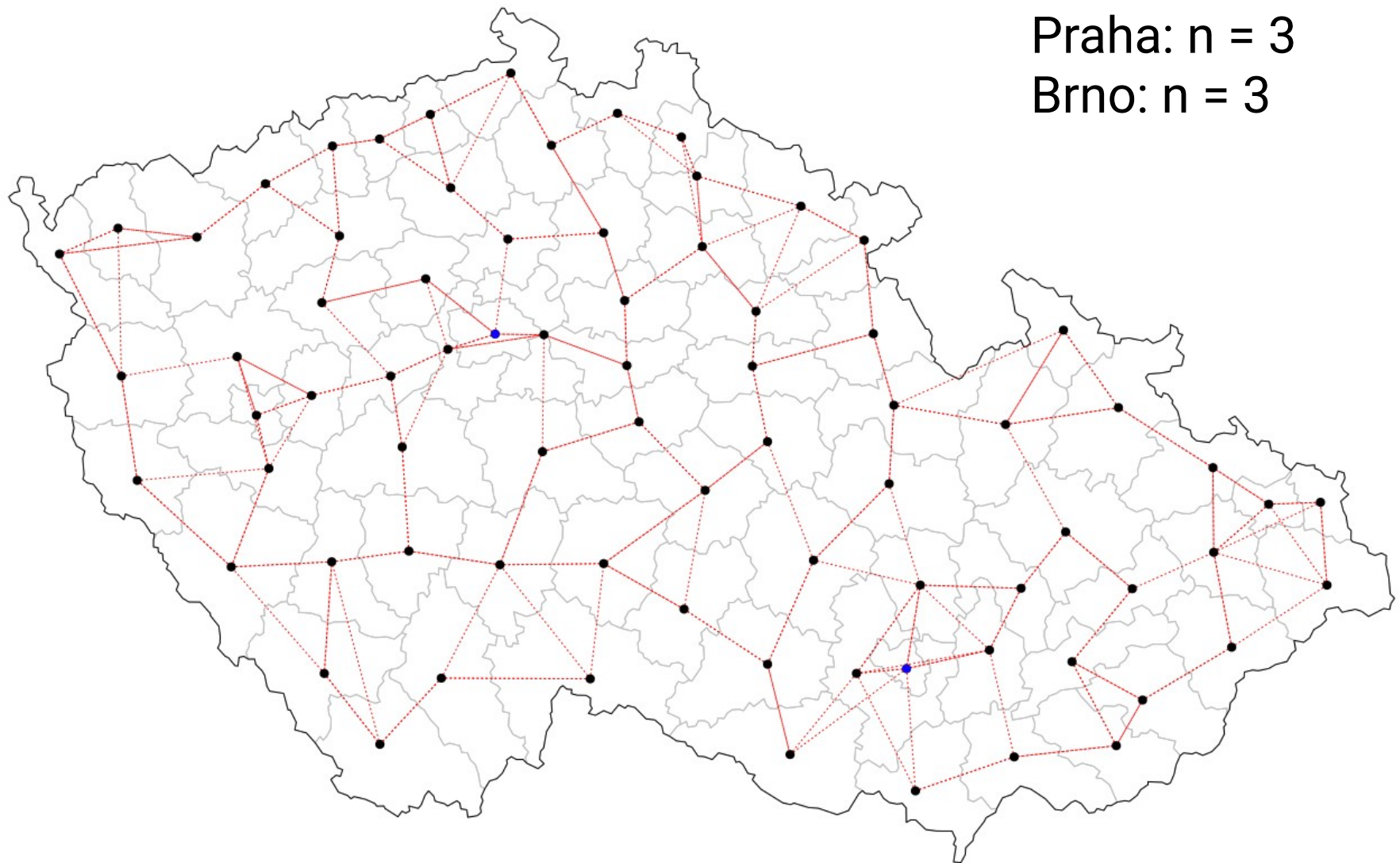
Řešený příklad

7-weight-matrices.R

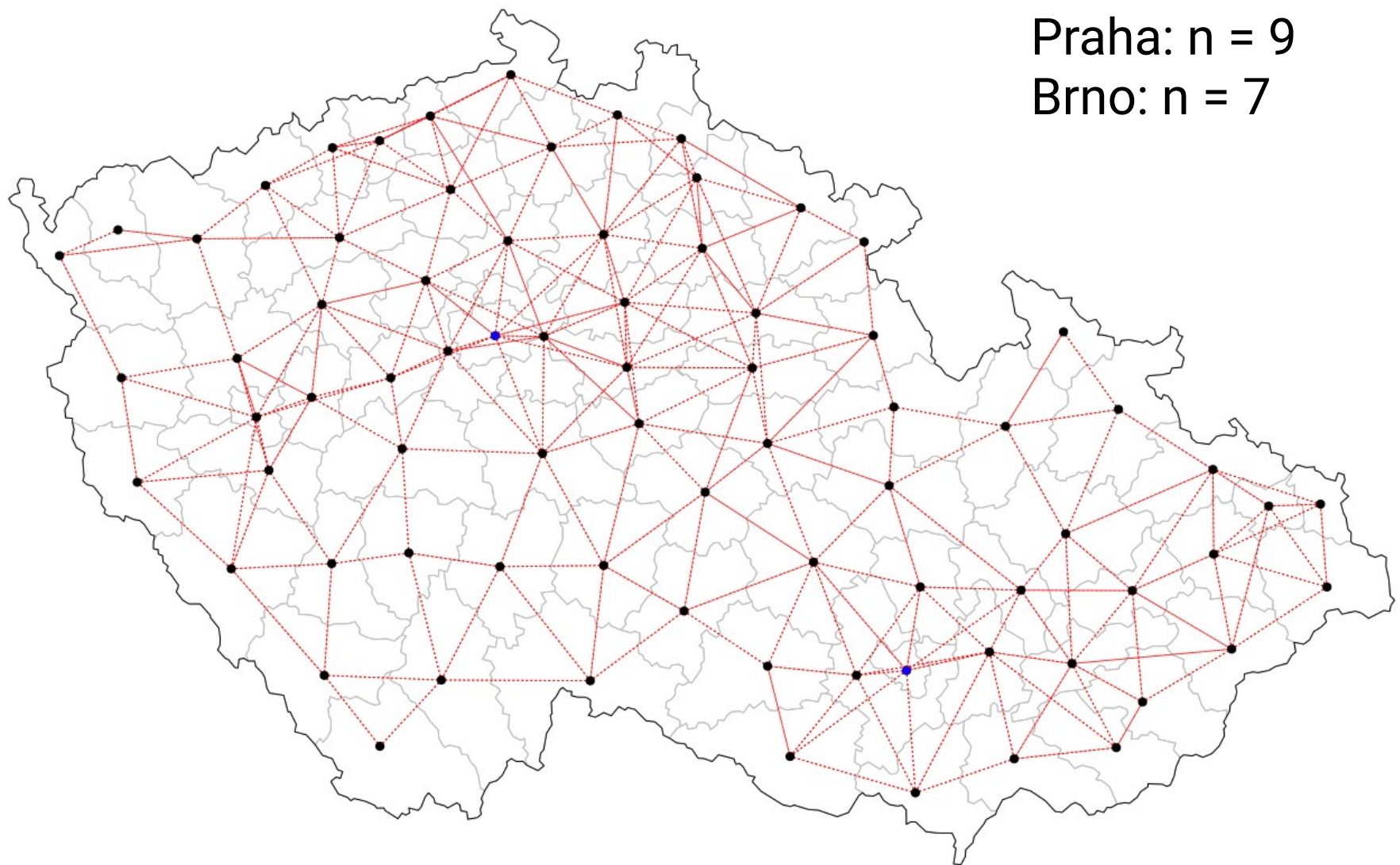
Sousedi věže



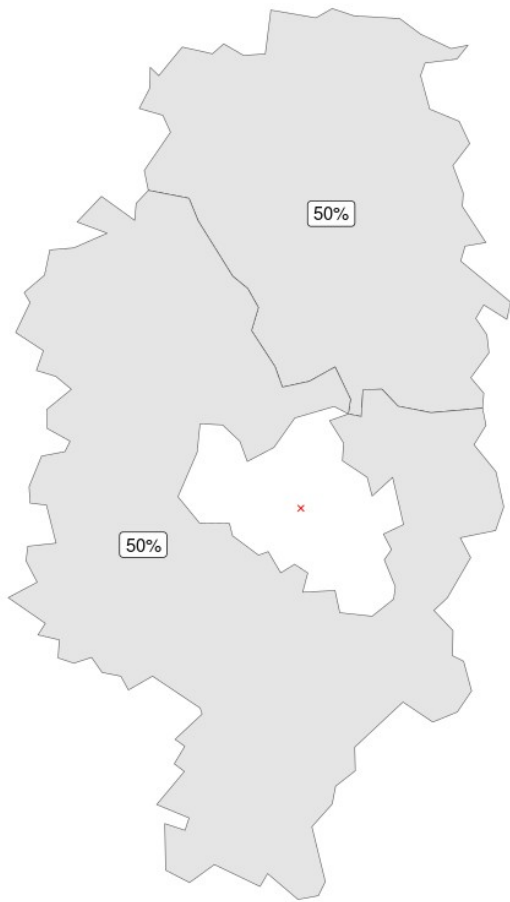
Sousedí 3NN



Sousedí 50 Km



Otázky nad maticemi



- Jsou shodné váhy všem fér?
- Stačí první soused?
- Kde skončit?
- Více sousedů s vahami často pomůže...

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)

Základní koncept Morana

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

- Základ = autokorelační koeficient
- Uplatňujeme váhy, nesoucí prostorovou interakci
- Váhy jsou analogem periody v časových řadách

Jak by vypadal svět...

- Kdyby všichni sousedi byli stejní?
- Moranovo $I = 1$
- Kdyby všichni sousedi byli odlišní?
- Moranovo $I = -1$
- Kdyby všichni sousedi byli náhodní?
- Moranovo $I = 0$

Rozvinutý koncept Morana

- Při splnění jistých předpokladůTM lze určit střední hodnotu & rozptyl statistiky I
- Potom lze přepočíst I na z-score, které má normální rozdělení

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

- Není-li praktické přijmout předpoklady normality, můžeme testovat metodou Monte Carlo = přeskládat hodnoty v matici sousedů mnohokrát *opravdu* náhodně, a porovnat histogram s napočtenou hodnotou

Technická implementace

- Test Moranovo I s metodou předpokladu normality (z-score)
`spdep::moran.test(x, w)`
- Test Moranovo I metodou Monte Carlo (zpřeházení sousedů)
`spdep::moran.mc(x, w, nsim)`

Řešený příklad

7-variace-na-moranovo-téma.R

Přístupy ke clusterům

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

Interpretace clusterů

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

Přístupy ke clusterům

- Lokální Moran I hodnoty
`spdep::localmoran()`
- Lokální Getis Ord G^* hodnoty
`spdep::localG()`

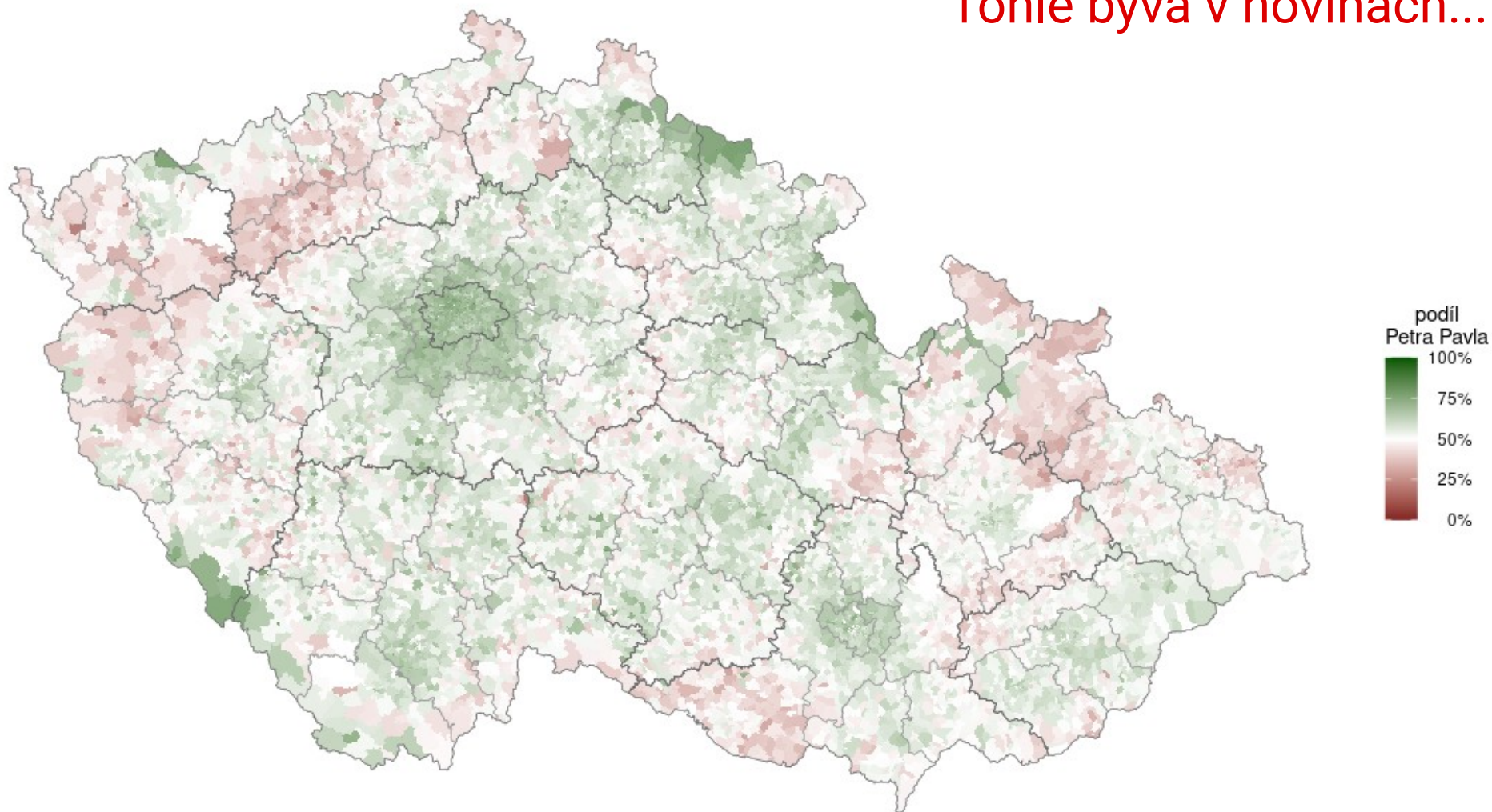
Prezidentské volby

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

Relativní hodnoty

Prezidentské volby 2023

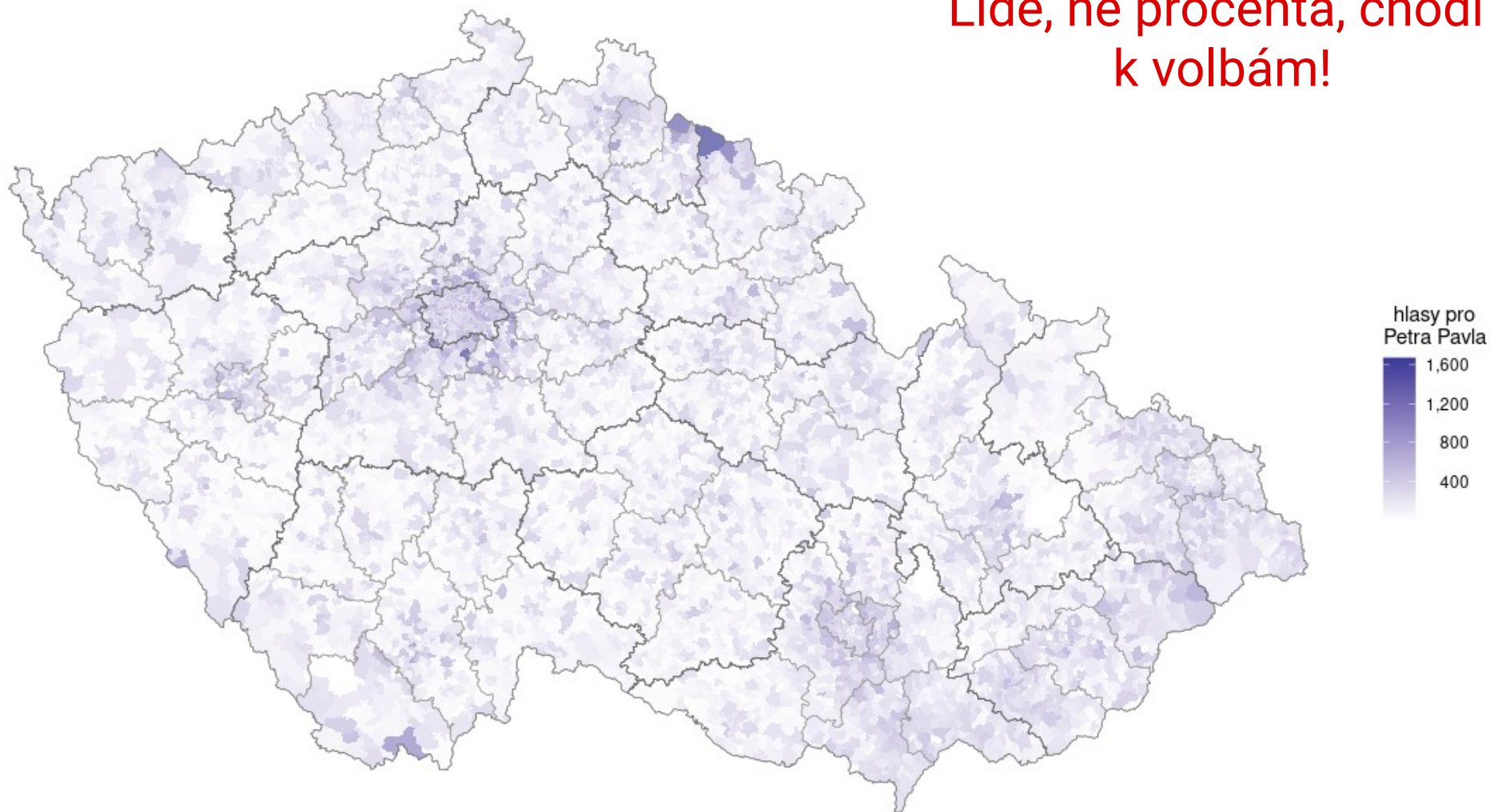
Tohle bývá v novinách...



Absolutní hodnoty

Prezidentské volby 2023

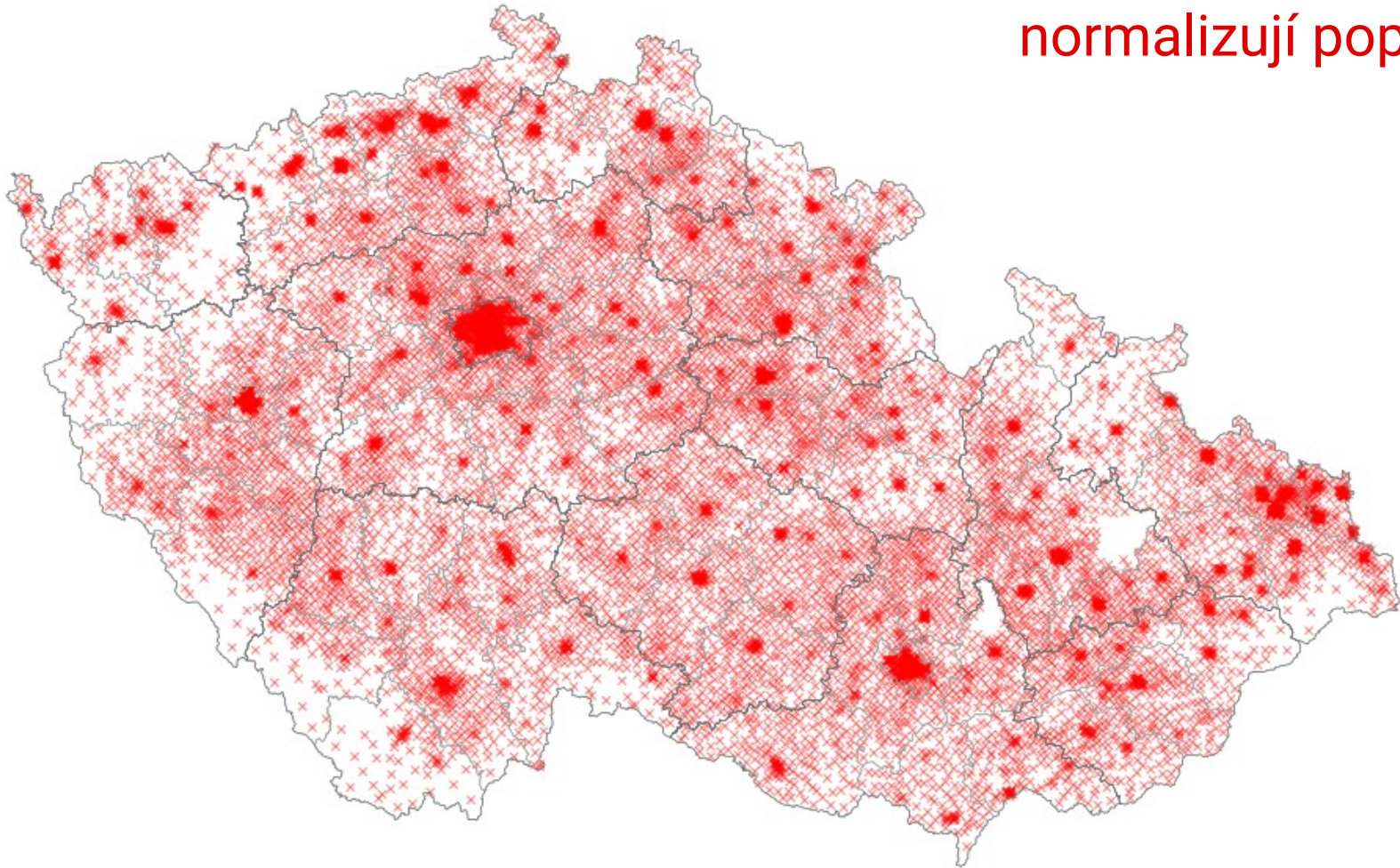
Lidé, ne procenta, chodí
k volbám!



Obvody a populace

Prezidentské volby 2023
středové body okrsků

Srovnatelně velké okrsky
normalizují populaci



Řešený příklad

7-digest-data-prez.R
7-okrsky-prez.R

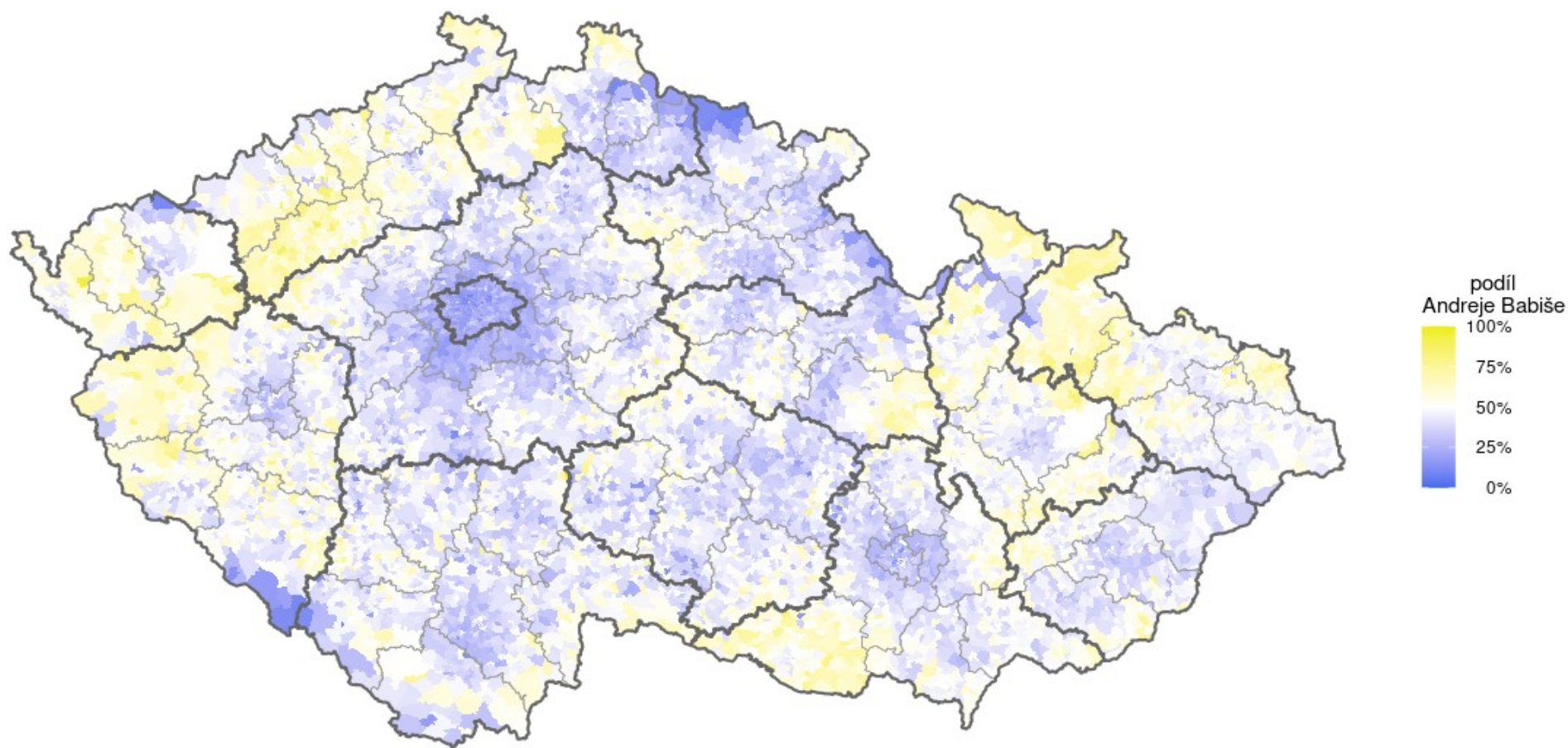
Řešený příklad

7-spdep-prez.R

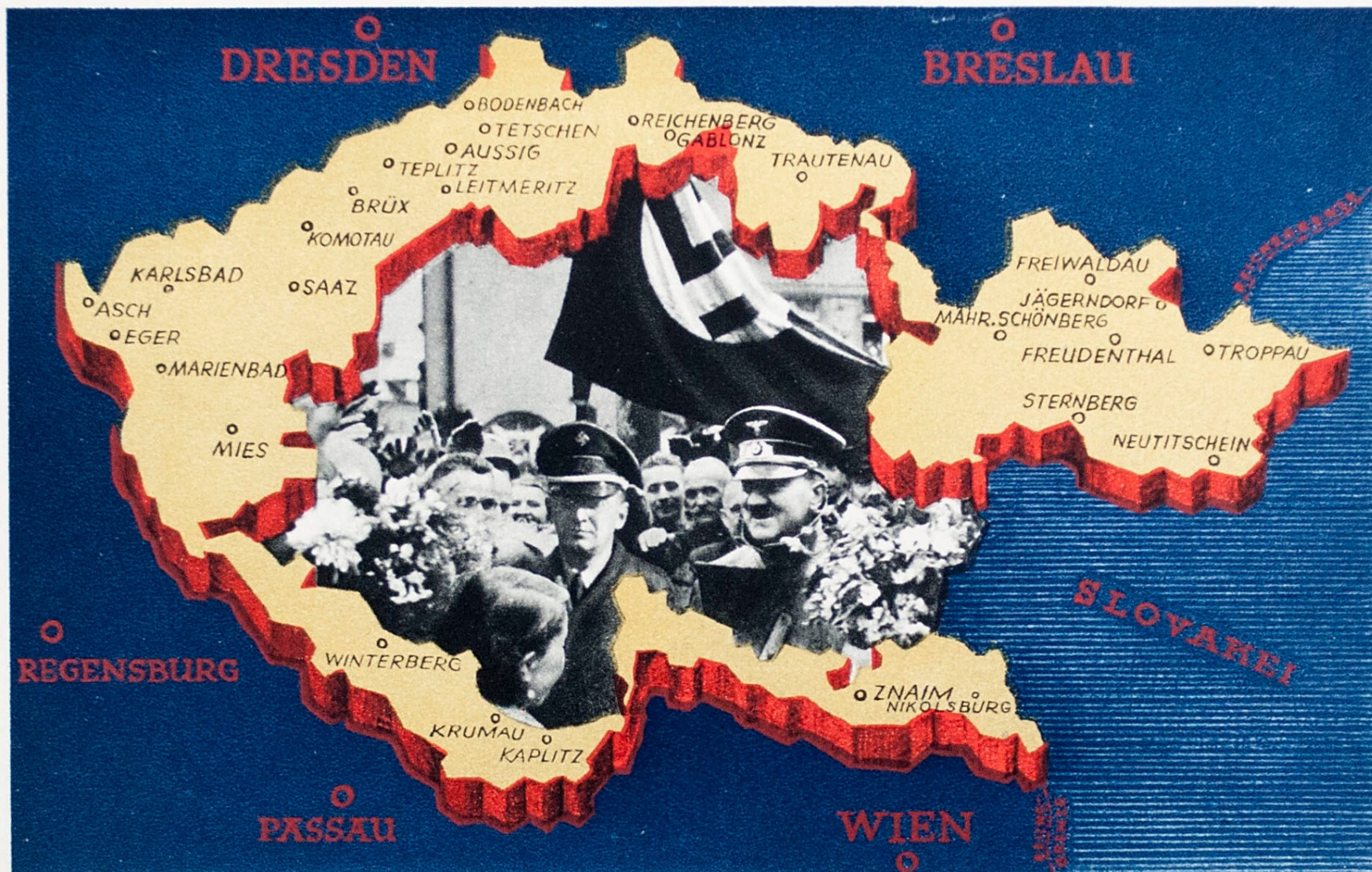
Příklad skryté veličiny

Prezidentské volby 2023

Připomíná vám tohle něco?



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

7-regrese-prez.R

Diskuze...