

O čem píší čeští sociologové a socioložky?

15 let Českého sociologického časopisu očima velkých jazykových modelů



KATEDRA SOCIOLOGIE
Filozofická fakulta
Univerzita Karlova

Aleš Vomáčka 10. 10. 2024

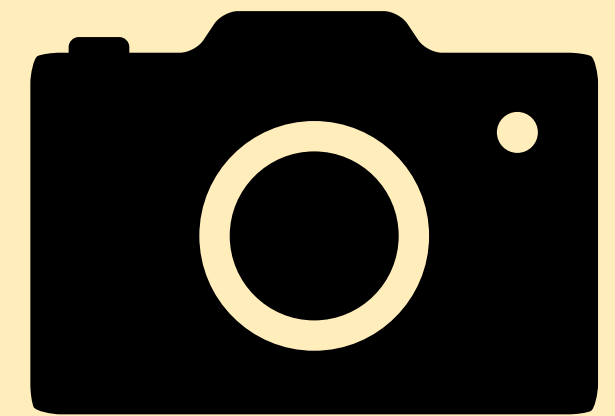
Strukturovaná data

0	1	0
1	0	1
0	1	0

Nestrukturovaná data

Lorem ipsum dolor sit amet,
consectetur adipiscing elit.
Suspendisse facilisis aliquam
lorem, vel mattis ligula
ullamcorper ac. Vivamus leo
lorem, pulvinar ac euismod ut,
feugiat iaculis sapien.

Aliquam mauris lacus, sodales
quis scelerisque quis, rutrum sed
mi. Sed accumsan varius tortor,
sit amet volutpat mauris
pellentesque id.



Dokumenty publikované mezi roky
2009 až 2024

414 abstraktů vědeckých studií
(30% všech dokumentů)

486 autorů/autorek
(v průměru 1.7 na článek)

1) Jaká témata se v abstraktech objevují?

2) Jak často se každé téma vyskytuje?

3) Souvisí výběr tématu s genderem autora/autorky?

1) Převést text na “embeddings”

2) Zredukovat počet dimenzí

3) Identifikovat témata

1) Převést text na “embeddings”

King - Man + Woman = Queen

2	4	6	8
5	2	9	6
1	7	5	8
8	4	0	3

6	5	4	0
6	4	6	1
2	3	5	8
1	3	9	2

0	2	4	1
2	2	6	2
3	8	5	3
1	0	7	6

-4	1	6	9
1	0	9	7
2	12	5	3
8	1	-2	7

GPT, Lamma, Claude, BERT,...

**Větší modely = kvalitnější
embeddings a větší kontextové
okno**

2) Zredukovat počet dimenzí

PCA, FA, UMAP,

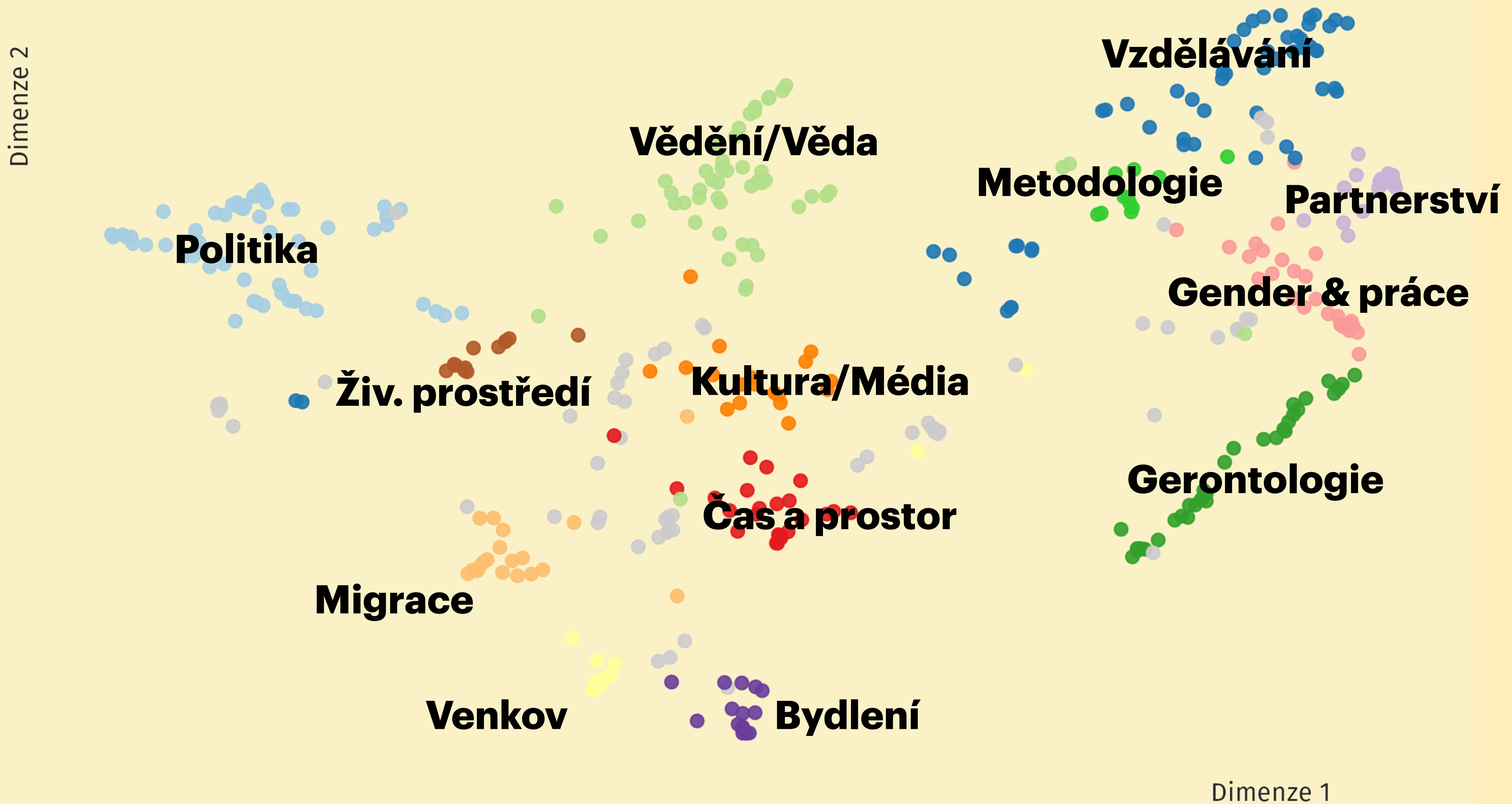
**Cíl obdobný jako u klasické
faktorové analýzy**

3) Identifikovat témata

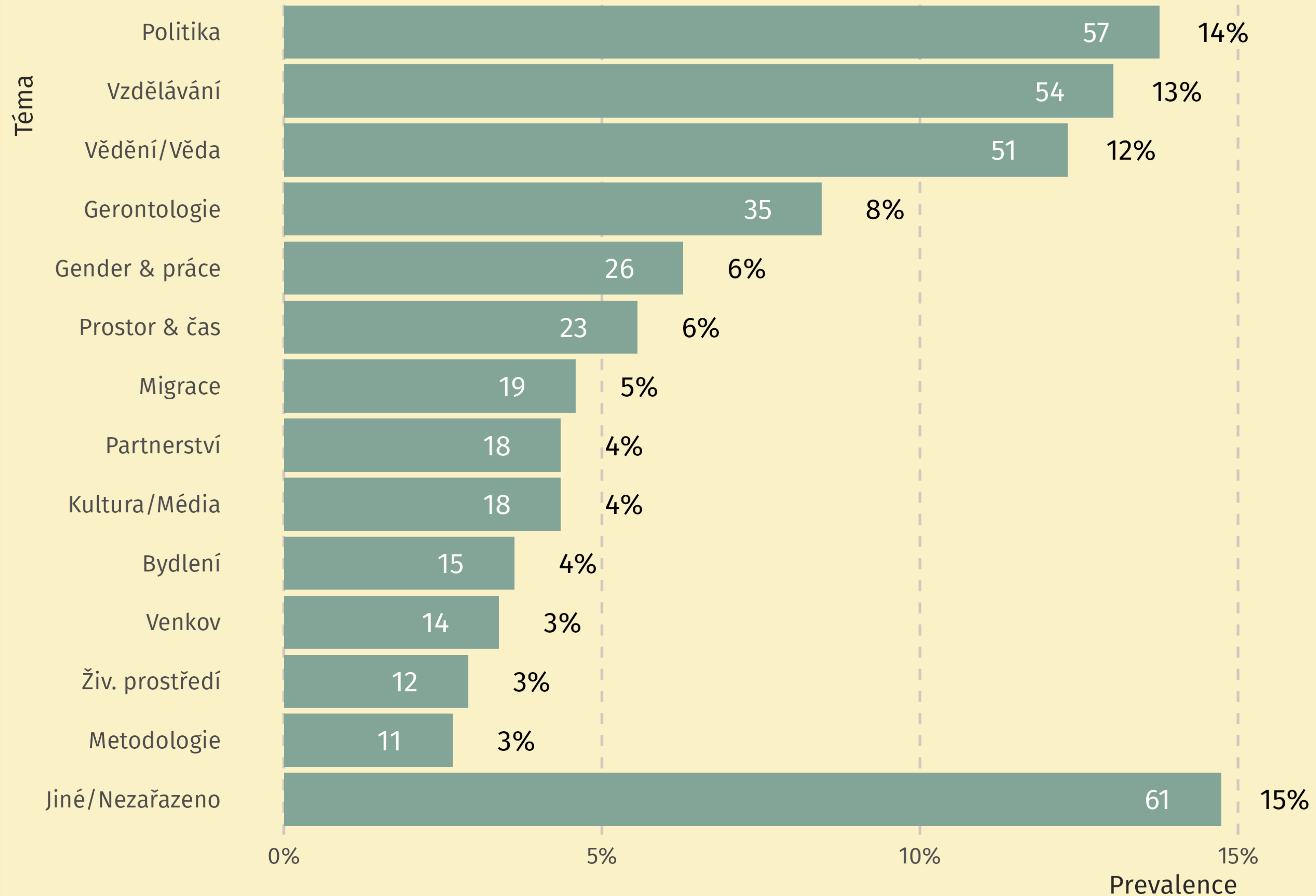
Hier. klast., K means, DBSCAN,...

**Výsledné klastry jsou
interpretované jako témata**

Mapa Českého sociologického časopisu 2009-2024



Politické sociologie zdaleka nejpopulárnější



n = 412. Relativní frekvence černě, absolutní bíle

Nic není dokonalé

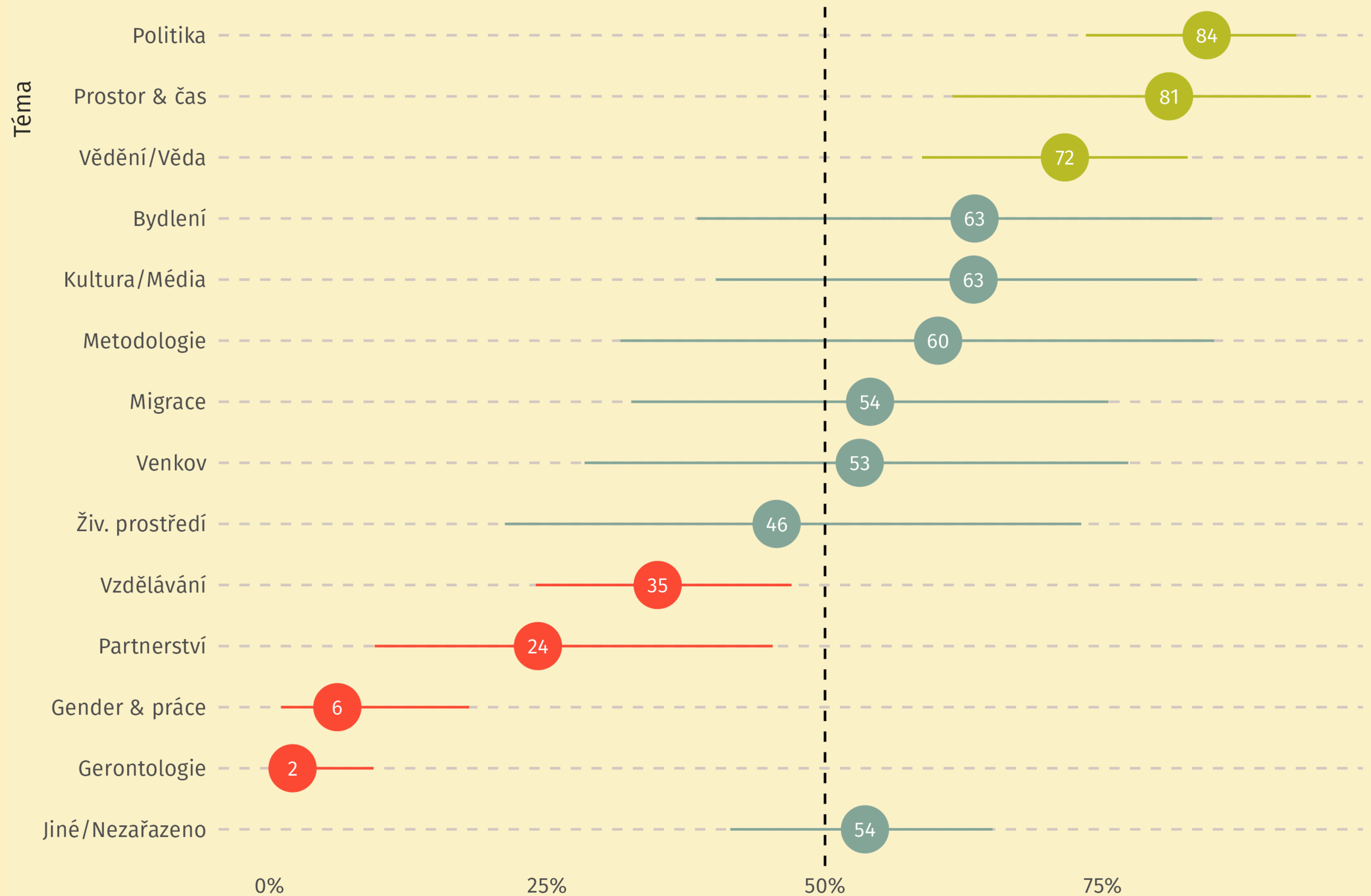
Trends in Divorce Acceptance and Its Correlates across European Countries

Vzdělávání?

Exponenciální modely náhodných grafů: modelování relačních mechanismů na případu sítě organizací zapojených v českém uhelném sektoru

Jiné/Nezařazeno

Politice dominují muži, genderu ženy



n = 412. Rozpětí představují 95% intervaly kredibility

A to je vše, díky za pozornost!



<https://github.com/alesvomacka/csr-scraping>

ales@vomacka.io

Politika

Strany na ústupu, lídři na vzestupu? Personalizace volebního chování v České republice

Gender a Práce

Neplodnost jeho a neplodnost její: Genderové aspekty asistované reprodukce

Gerontologie

Sourozenectví v pozdním věku: Příspěvek k teoretické diskusi

Vzdělávání

Vliv postojů učitelů na výsledky žáků

Bydlení

Private Rental Housing in the Czech Republic: Growth and...?

Venkov

Endogenní rozvojové potenciály malých venkovských obcí

Migrace

Transnacionalismus a stálost návratu v arménské návratové migraci

Kultura/Média

Discourses of Thrift and Consumer Reasonability in Czech State-Socialist Society

Čas a Prostor

"Předevčírem, nebo kdy to bylo?": Temporalita třídy nejchudších

Vědění/Věda

K rekonstrukci "kapitalismu" u autorů otevřeného marxismu

Metodologie

Measurement Invariance of the SQWLi Instrument Over Time

Životní prostředí

Environmentálně orientované motivace a potenciál zklamání