Data – zpětná vazba

- Když máte mnohem více/méně řádků, než je obcí v obvodu, tak je někde problém
- Když se často opakují prázdné řádky, tak je někde problém
  - Typická příčina = začátek od jiných než senátních voleb
  - řešení -> seřadit podle voličů v seznamu
- Když se opakují čísla obcí, tak je někde problém
  - Příčiny:
    - stejné názvy sloupců ve více souborech
    - Zapomenutá okrsková data

Názvy sloupců

- Co možná nejkratší
- Bez diakritiky
- Bez teček čárek a mezer
- Nesmí začínat číslem (spss přidá @, qgis pak bude mít problém)

- Např. „Zahradaníček proc.hlasů v 1. kole 2018“ -> Zahr_18_1p

Pokročilé postupy

- Změny obcí (Libhošť, Krhová, Poličná, Ladná, Polná na Šumavě, Město *Libavá, Kozlov a Luboměř pod Strážnou*)
- Možnost využít okrsková data původní obce před vznikem nové obce
- Příklad: https://www.volby.cz/pls/senat/se211?xjazyk=CZ&xdatum=20121012&xobvod=77&xobec=545058
- Nutné vědět, co znamenají čísla volebních okrsků

Data

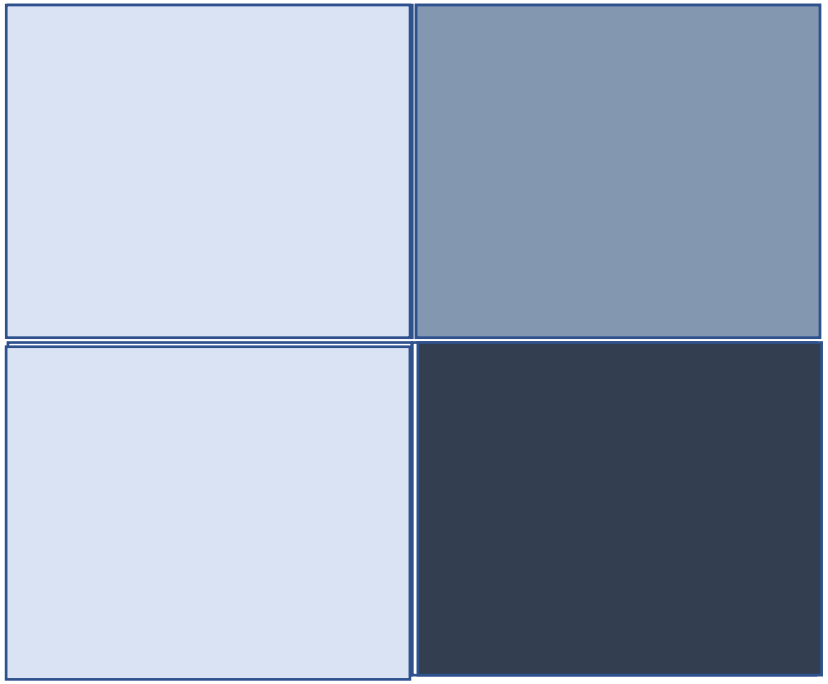
Úvodem

- Prostorová analýza x analýza v prostoru
  - V **politologii** se prostorové hlasovaní (spatial voting) a prostorová analýza voleb (spatial analysis) týká myšleného politického prostoru
  - V **geografii** se prostorová analýza týká fyzického prostoru
- Kvantitativní = potřeba vhodných dat
- V tradičním pojetí prostorové analýzy se neuplatňují jiné než kvantitativní metody

Povaha dat

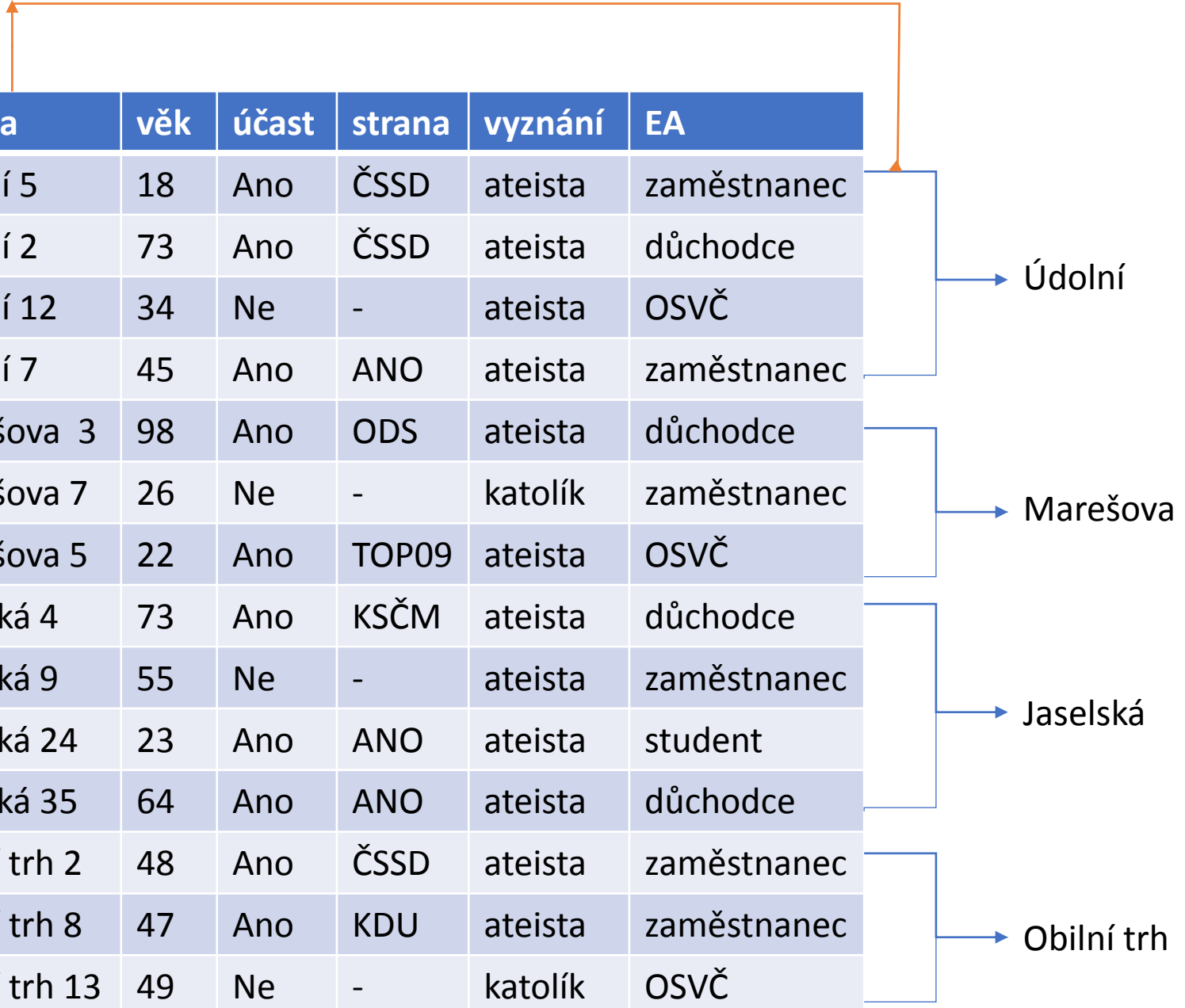
- Agregovaná data
- Kardinální
- Prostorové
  a časové zařazení

Princip vzniku agregovaných dat



- Různě vysocí lidé
- V několika místnostech
- Agregace dat
  - Jen jeden údaj za místnost
  - Různé situace mohou vést ke stejnému výsledku

Jméno	Adresa	věk	účast	strana	vyznání	EA
Karel J.	Údolní 5	18	Ano	ČSSD	ateista	zaměstnanec
Jana B.	Údolní 2	73	Ano	ČSSD	ateista	důchodce
Jiří K.	Údolní 12	34	Ne	-	ateista	OSVČ
Květa D.	Údolní 7	45	Ano	ANO	ateista	zaměstnanec
Tomáš V.	Marešova 3	98	Ano	ODS	ateista	důchodce
Marie H.	Marešova 7	26	Ne	-	katolík	zaměstnanec
Jan Z.	Marešova 5	22	Ano	TOP09	ateista	OSVČ
Eva A.	Jaselská 4	73	Ano	KSČM	ateista	důchodce
Zdeněk C.	Jaselská 9	55	Ne	-	ateista	zaměstnanec
Pavel N.	Jaselská 24	23	Ano	ANO	ateista	student
Anna R.	Jaselská 35	64	Ano	ANO	ateista	důchodce
Emil M.	Obilní trh 2	48	Ano	ČSSD	ateista	zaměstnanec
Lucie S.	Obilní trh 8	47	Ano	KDU	ateista	zaměstnanec
Milan T.	Obilní trh 13	49	Ne	-	katolík	OSVČ



Jméno	Adresa	věk	účast	strana	vyznání	EA
Karel J.	Údolní 5	18	Ano	ČSSD	ateista	zaměstnanec
Jana B.	Údolní 2	73	Ano	ČSSD	ateista	důchodce
Jiří K.	Údolní 12	34	Ne	-	ateista	OSVČ
Květa D.	Údolní 7	45	Ano	ANO	ateista	zaměstnanec

(18+73+34+45)/4

3xAno/4

ulice	Průměrný věk	účast	ČSSD	ANO	KSČM	ODS	TOP	KDU	kato líci	zaměs tanci	Poč et
Údolní	42,5	75	66,6	33,3	0	0	0	0	0	50	4
Marešova	48	66,6	0	0	0	50	50	0	33,3	33,3	3
Jaselská	50,1	75	0	66,6	33,3	0	0	0	0	25	4
Obilní trh	48	66,6	50	0	0	0	0	50	33,3	66,6	2

EKOLOGICKÁ CHYBA

- V procesu agregace se ztrácí informace
- Nelze hovořit o pozorovaných vztazích jako o platných pro individuální voliče
  - V našem případě: tam kde je nějaký katolík získala hlas KDU nebo TOP09. Přitom ale oba katolíci nevolili.
  - V reálném světě nejsme schopni věrohodně z agregovaných dat věrohodně rekonstruovat individuální vztahy
  - Rozdíl akademický výzkum X práce v kampani

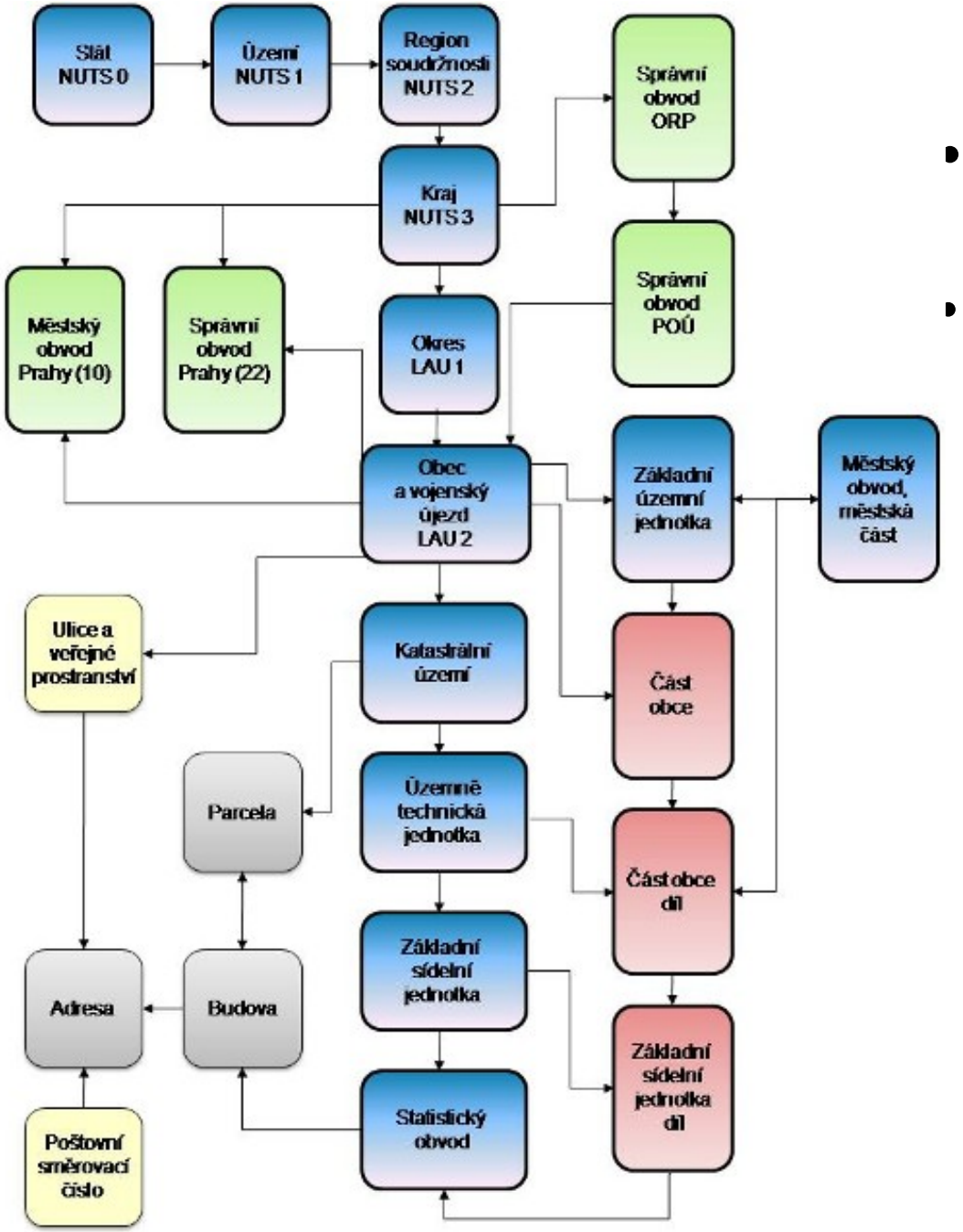
Specifika prostorových dat

- Autokorelace (více seminář 3)
  - „vše souvisí se vším, a co si je blíž, to spolu souvisí více“
  - Porušení předpokladu o nezávislosti pozorování
- Nestacionarita (více seminář 4)
  - Volební chování jedné společenské skupiny se může v prostoru lišit (katolíci ve Valašských Kloubocích x katolíci v severních Čechách, podnikatelé v Praze x podnikatelé na Svitavsku)

Velikost polygonu

- Malý region = vysoká homogenita/
  vysoký „šum“
- Velký region = nízká homogenita/
  nízký šum

  - Funkční x administrativní region



- http://apl.czso.cz/iso4/cisel.jsp
- Viz http://www.cuzk.cz/Uvod/Produkty-a-sluzby/RUIAN/2-Poskytovani-udaju-RUIAN-ISUI-VDP/Ciselniky-ISUI/Nizsi-uzemni-prvky-a-uzemne-evidencni-jednotky.aspx

Zdroj: google.maps.com

Rozdíly mezi měřítky

	nezaměstnanost		vš		Počet obyvatel		OSVČ		
	min	max	min	max	min	max	min	max	
Brno	9,4		20,6		385913		14,19		1
Městské části	5,8	11,9	11,1	32,5	577	64316	9,2	18,3	29
„Čtvrtě“	0	23	0	39	5	26781	6	75	59
ZSJ	0	100	0	75	0	12836	0	100	264

- kvůli značným rozdílům ve velikosti jednotek je obvykle vhodné používat váhy

Důsledky „měření“ (sběru dat)

- V socio-ekonomické analýze obvykle není problém s chybějícími daty za místa
- Problém s chybějícími daty pro čas (mnoho údajů je zjišťováno jen z cenzu)

- Bojkot sčítání (např. Řekové v Albánii, Albánci v Makedonii a Srbsku, …)
  - V ČR otázka víry a vyznání v roce 2011

Data za obce dostupná každoročně (nebo častěji)

- Počet obyvatel
- Věkové složení
- Pohyb obyvatelstva (narození/zemřelí, přistěhovalí/vystěhovalí)
- Bytová výstavba
- Nezaměstnanost (měsíčně)
- Rozpočty obcí
- Školská statistika

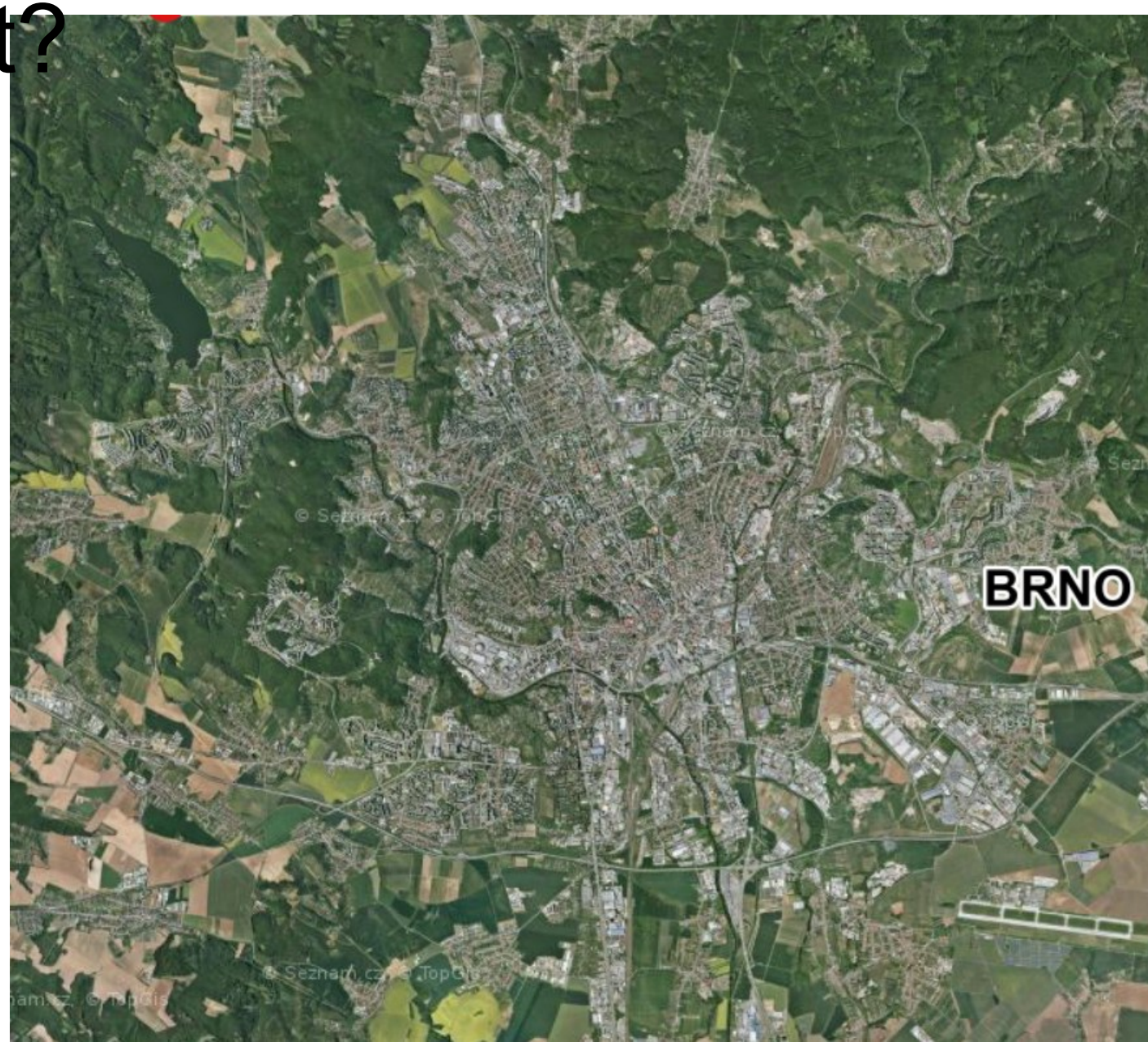
Shrnutí

- Agregovaná data
  - Nebezpečí ekologické chyby
- Kardinální proměnné
  - Možnosti pro využití řady statistických nástrojů
- Prostorová data
  - Narušení obvyklých předpokladů
  - Otázka měřítkové úrovně
  - Otázka spolehlivosti dat

Vážení dat

Co je to vážení dat?

- Jihomoravský kraj
  - Nejmenší obec: Říkonín, 33 obyvatel
  - Největší obec: Brno, cca 380 000 obyv.
- Bez vážení dat znamená procento voličů v Řikoníně totéž co procento voličů v Brně
- Pro volební výsledek to ale není totéž



Výpočet váhy

- Možnost vážit populací obce (buďto počet obyvatel, nebo v našem případě lépe počet voličů)
  - Zkresluje hodnoty standardní chyby (ta nás sice moc nezajímá, ale pro jistotu)
- Vhodný způsob: podíl obce na voličích/obyvatelstvu obvodu
  - suma sloupce voličů nebo obyvatelstva
  - vydělení počtu voličů-obyvatel obce sumou za celý obvod
  - Součet je 1
  - Vynásobení počtem obcí v kraji (např. 500 obcí)
    v excelu: vaha = a2/a501*500, kde a501 je suma voličů
    v spss: transform -> compute vaha = volici08/suma volici*500

  Počet případů ve vážené analýze odpovídá reálnému počtu obcí

Použití váhy

- Data – weight cases (úplně dole)
- Weight cases by – do pole Frequency variable vložte proměnnou vaha - Ok

Důsledky použit vah

- Spočítané charakteristiky jsou blíže „realitě“
- Situace ve městě s 10 000 obyvateli má pro výsledky stejný dopad jako situace v 10 obcích s 1000 obyvateli nebo ve 100 obcích se 100 obyvateli
- Bez vah analýzy odráží spíše situaci malých obcích, kterých je sice mnoho, ale vzhledem k počtu voličů nemusí být pro podporu strany důležité

Popis podpory pomocí deskriptivních statistik

Popis volební podpory

- Kde?
- Jak?

Proč?

- Účelem deskriptivní statistiky je zjistit vlastnosti proměnné
  - (Porozumět používaným datům)
- Vlastnosti proměnné mají důsledek pro další analýzy
- A pro interpretaci výsledků analýz

- Zisk zjednodušené informace o volební podpoře stran/kandidátů

Jakou podporu mají kandidáti v obvodu?

Známe od nepaměti

- Minimum a maximum
  - Dávají velmi hrubou představu, v jakém rozmezí se podpora strany pohybuje
  - Mohou indikovat případné chyby ve výpočtu

Známe z loňska

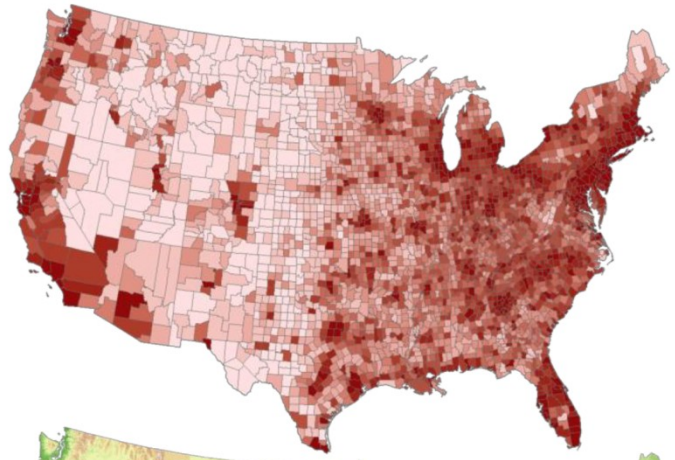
- Průměr - Hledáme hodnotu, která nejlépe reprezentuje proměnnou
  - Samotná střední hodnota poskytuje značně redukovanou (a zkreslenou, pokud jsou v datech odlehlé případy) představu o vlastnostech proměnné
  - Vhodné srovnat s mediánem
- Míra variability doplňuje informaci, jak dobře střední hodnota reprezentuje všechny případy
  - Ukazuje, jak moc se mezi sebou liší hodnoty proměnné
- Kvantily
  - Ukazují mezní hodnoty, ve kterých se pohybuje určité procento případů

Jak je volební podpora stran ne/koncentrovaná?

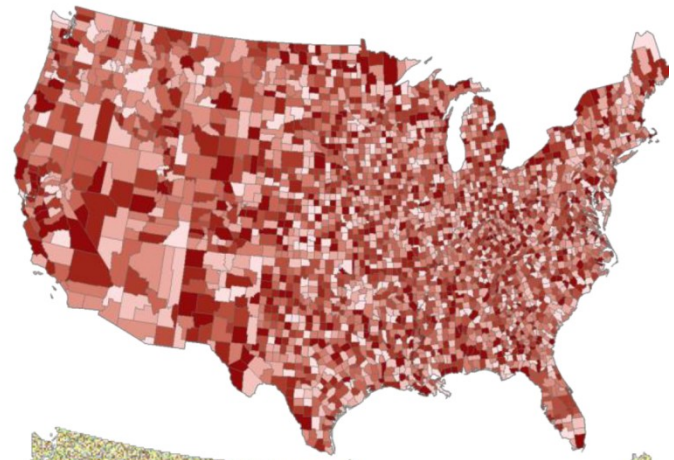
- Ze statistického hlediska
- Variační koeficient
- Giniho koeficient

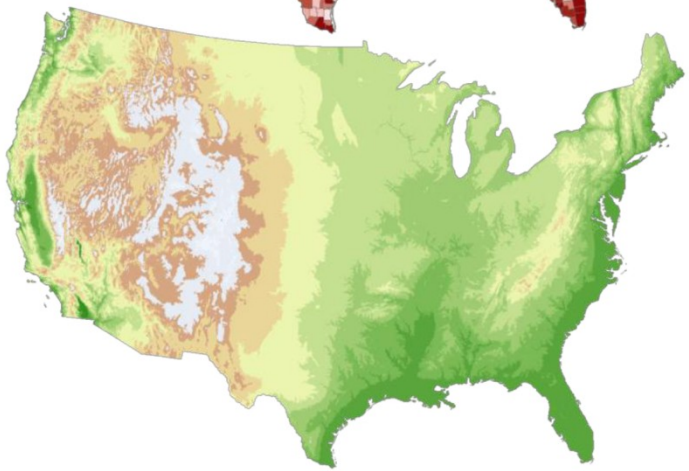
- Z geografického hlediska
- Moranovo I
- Lisa

If features were randomly distributed ...

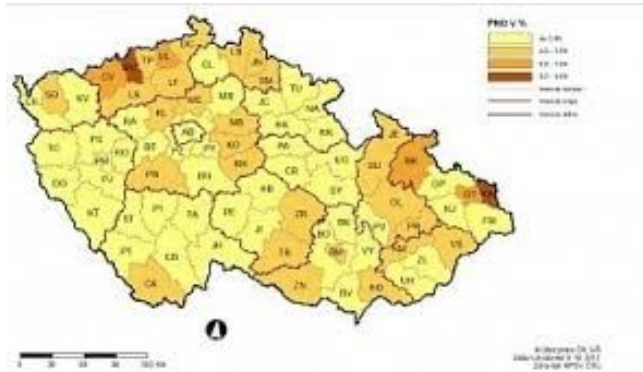


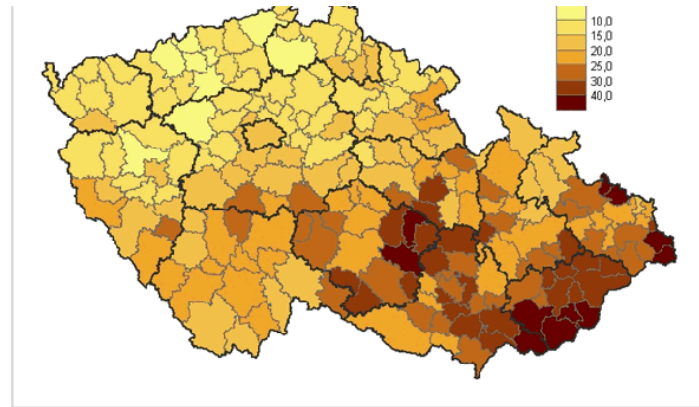
... population density map of the US would look like this





... elevation map of the US would look like this





Variační koeficient

- =směrodatná odchylka/průměr
- V prostorové analýze využíván jako míra koncentrace
- Neznamená ale míru územní koncentrace!!!
- Nebere v potaz rozložení hodnot v prostoru
- 0 – velmi malé rozdíly v hodnotách proměnné
- Nemá pevnou horní hranici
  - Není vhodné interpretovat v procentech
  - Vhodný referenční bod pro interpretaci jsou hodnoty var. Koef. ostatních kandidátů

Giniho koeficient

- Rozdíl skutečného a ideálně rovnoměrného rozložení
- Stále nebere v potaz uzemní koncentraci
  - Ale lépe zohledňuje velikost obcí
- Není obsažen v spss ani excelu
  - Postup v manuálu v materiálech

Popis vývoje volební podpory

Jak se liší volební podpora mezi poslaneckými a krajskými volbami?

- Rozdíly v popisných statistikách
- korelace

- Mapa „volební úspěšnosti“
- Území stabilní volební podpory

Území volební podpory

- Přehlednost, zohlednění populační velikosti
- Vhodné pro sledování delšího časového úseku
- Nevhodné pro lokální strany (např. SMK na Slovensku)
- Nevhodné v nestabilních systémech

- Doplňující indikátory
  - Míra úspěšnosti: kolikrát je podpora strany vyšší v (jádru) území volební podpory oproti zbytku území
  - Výše volební podpory: jaké procento hlasů v obci, která má z území nejnižší podporu

Území stabilní volební podpory

- Strany s koncentrovanou stabilní podporou X strany s nekoncentrovanou stabilní podporou
- Výpočet procenta hlasů přítomných v ÚSVP v každých volbách
- Zbytečné pro krátký časový usek

Bazické a řetězové indexy

- Řetězový index
  - Volby 2017/volby 2013; volby2013/volby2010 …
  - Ukazuje postupný vývoj
- Bazický index
  - Volby 2017/volby2006, volby2013/volby2006 …
  - Ukazuje změnu stav oproti stanovenému základu

  - Index volební úspěšnosti
    - =Kraj 2018/poslanecké volby 2017 * 100
  - Index volební stability
    - kraj 2018/ kraj 2014* 100

Korelační analýza

- Co se dozvíme z výsledku korelační analýzy?
- Proč může korelační analýza posloužit k analýze vývoje volební podpory?
- Jaký typ souvislosti sleduje?
- Co z korelační analýzy zjistíme o příčinách změny volební podpory?

Korelační analýza

- Ukazuje lineární souvislost mezi dvěma proměnnými
- V prostorové analýze ukazuje:
  - Jak stabilní je rozložení podpory strany mezi 2 volbami
  - Jak se liší rozložení podpory dvou stran v jedněch volbách
- O příčinách se nedozvíme nic zásadního
  - Můžeme zjistit,
  - Jestli došlo ke změně rozložení podpory
  - jestli byl pokles/nárůst podpory rovnoměrně rozprostřený nebo koncentrovaný

Korelační analýza

- Pokud se korelační koeficient blíží 1 pak se rozložení podpory nezměnilo
  - Stále mohlo ale dojít k poklesu nebo nárůstu podpory
- Je míra souvislosti stejná s výsledky v různých typech voleb?
  - Pokud se výsledky strany v krajských a parlamentních volbách velmi liší, možná je podpora dána jinými faktory (např. efekt lídra?)