

# Prostorová analýza voleb seminář 1

# Organizace seminářů

- Program v sylabu je orientační
  - Pokud nebude v předstihu uvedeno jinak, literaturu k seminářům čtete PŘED seminářem
- Dnes: data, příprava dat, ověření úrovně schopností v oblasti statistické analýzy
- Příště: vytváření mapy
- Přes příště: popisné statistiky, korelace
- Nakonec: „prostorové“ metody
- Úplně nakonec: shrnutí/opakování ???

# Semináře

- K čemu jsou:
  - Jak zjistit z agregovaných dat užitečné informace o volebním chování
    - Prakticky využitelné zejména pro práci s volbami, ke kterým se nedělají průzkumy veřejného mínění
  - Rozvoj dovedností v oblasti statistických metod
    - Počítáme s tím, že něco umíte z podzimního kurzu
  - Seznámení s „kartografickými“ metodami
  - Zpracování analýz pro seminární práci
  - OBSAH SEMINÁŘŮ ANI LITERATURY URČENÉ K SEMINÁŘŮM NEBUDE SOUČÁSTÍ ZKOUŠKY
    - Osvojení si dovedností a znalostí bude otestováno seminární prací
    - Literaturu rozhodně čtěte

# Seminární práce

- Cíl: popsat a vysvětlit rozdíly v podpoře kandidátů v senátních volbách
- Jednotlivé kroky
  - Popis volební podpory pomocí map
  - Popis volební podpory pomocí deskriptivních statistik
  - Popis vývoje volební podpory pomocí analýzy souvislosti s předchozími volbami
  - Vysvětlení rozdílů v podpoře kandidátů v obcích pomocí regresní analýzy
- PRO ZPRACOVÁNÍ PRÁCE BUDOU DŮLEŽITÉ JAK INFORMACE Z TOHOTO KURZU, TAK Z PODZIMNÍHO KURZU O KVANTITATIVNÍCH METODÁCH

# Seminární práce 2.

- Jednotlivé kroky budou konkrétněji zadány na dalších seminářích
- Finální práce bude kompilací úkolů + úvod, závěr, literatura, propojení jednotlivých částí vysvětlujícími pasážemi
- V hodnocení bude také zohledněno, jakým způsobem byly vyřešeny výtky sdělené v hodnocení úkolů

# úkoly

- Za účast na seminářích a za odevzdání všech úkolů získáte 10 bodů
- Neodevzdání některého z úkolů = 0 bodů
- Úkoly samotné nebudou nijak hodnoceny, jen vám přinesou zpětnou vazbu
  - Odevzdaný úkol, který neřeší zadání, bude považován za neodevzdaný
- Neúčast na semináři je možná, ale přístup k materiálům k danému semináři budou mít jen zúčastnění
- Termín odevzdání bude sdělen vždy se zadáním úkolu (obvykle půjde o pondělní půlnoc před středečním seminářem)
- Zadání úkolu bude zveřejněno vždy na konci semináře
- Odevzdání úkolů vede k vyšší šanci vytvoření kvalitní seminární práce (a tedy k vyšší šanci na dobré bodové hodnocení)

Data

# Úvodem

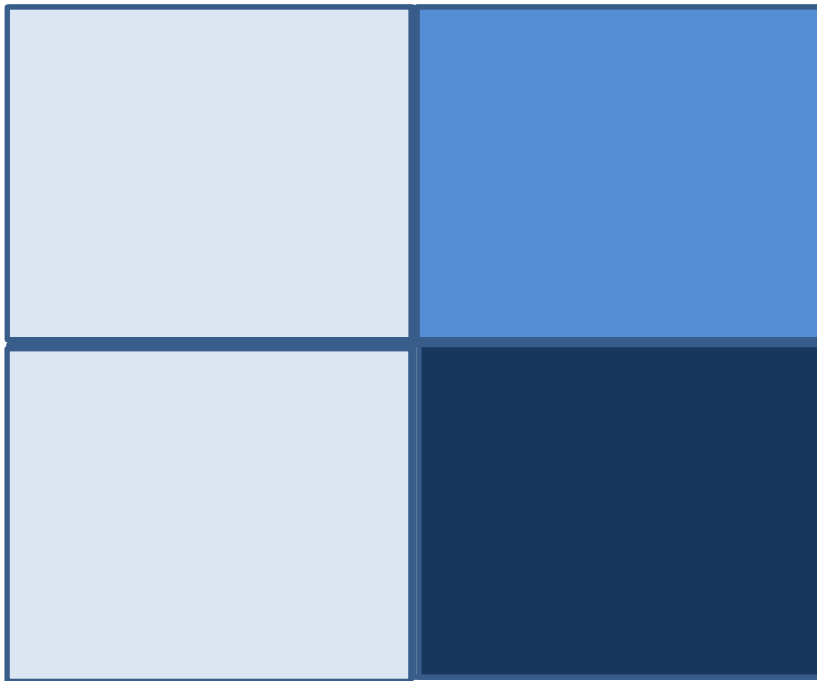
- Prostorová analýza x analýza v prostoru
  - V **politologii** se prostorové hlasování (spatial voting) a prostorová analýza voleb (spatial analysis) týká myšleného politického prostoru
  - V **geografii** se prostorová analýza týká fyzického prostoru
- Kvantitativní = potřeba vhodných dat
- V tradičním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



# Povaha dat

- Agregovaná data
- Kardinální
- Prostorové  
a časové zařazení

# Princip vzniku agregovaných dat



- Různě vysokí 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

Údolní

Marešova

Jaselská

Obilní trh

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

$$(18+73+34+45)/4$$

$$3 \times \text{Ano} / 4$$

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

# EKOLOGICKÁ CHYBA

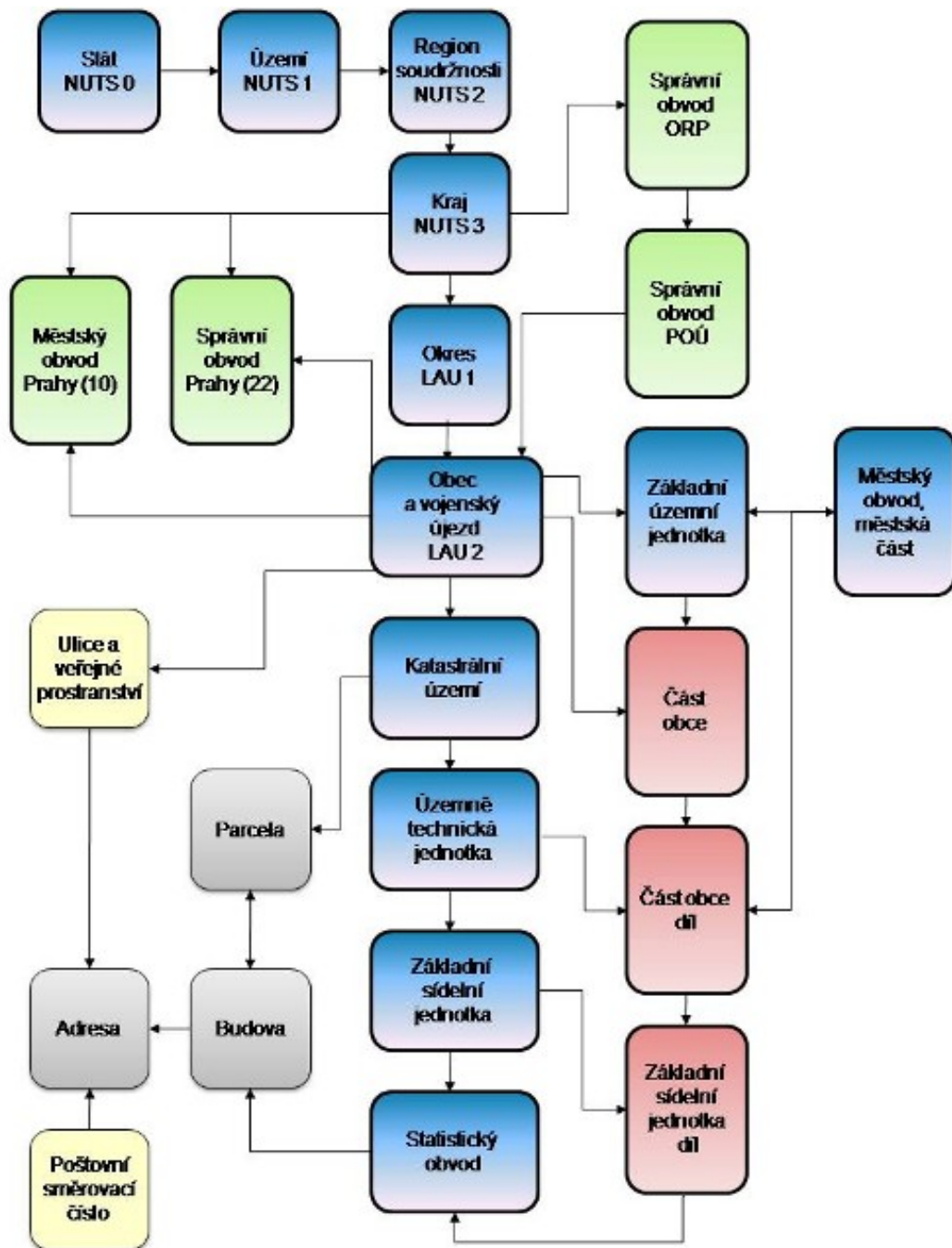
- V procesu agregace se ztrácí informace
- Nelze hovořit o pozorovaných vztazích jako o platných pro individuální voliče
  - V našem případě: tam kde je nějaký katolík získala hlas KDU nebo TOP09. Přitom ale oba katolíci nevolili.
  - V reálném světě nejsme schopni věrohodně z agregovaných dat věrohodně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

# 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 Klouboucích 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/irso4/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

# 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

Jak nachystat data?

# Data – volební výsledky

- 2 možnosti jejich zisku
  - Volby.cz
  - Otevřená data
    - (žádost na ČSÚ – v případě dat, která nejsou v otevřených datech)
    - Data za obce
    - S okrsky je těžká práce
    - Je známé jen aktuální vymezení okrsků
    - <https://volby.tmapy.cz/>

# Volby.cz

- Otravná manuální práce
- Možnost obejít automatizovaným stahováním (tzv. webscrape)
- Skript lze napsat v VBA/pythonu (excel) nebo v R
  - To se tady učit nebudeme
  - Pokud chcete stahovat data z volby.cz nebo z jiných serverů se systematicky uspořádanými tabulkami, pak se tato schopnost velmi hodí
  - <http://analystcave.com/web-scraping-tutorial/>
  - <http://analystcave.com/excel-tools/excel-scrape-html-add/>
  - <https://www.promptcloud.com/blog/how-to-use-excel-to-scrape-websites>



# Otevřená data

- Součást serveru volby.cz
- <http://volby.cz/opendata/opendata.htm>
- Obsahuje systematizovanou informaci o kandidujících subjektech (**registry**) a volebních výsledcích (**okrsková data**)
- Stažení zazipované složky



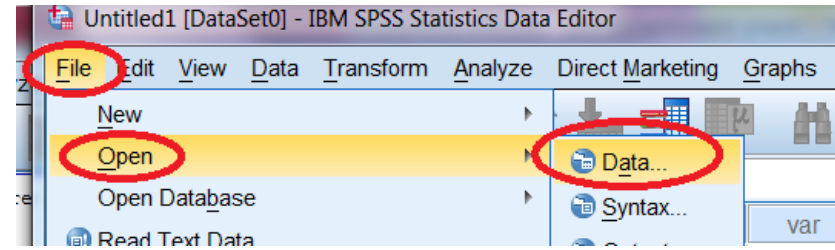
# Úprava excelového souboru

- Vyfiltrovat svůj obvod
  - Data – filtr – obvod – číslo obvodu
- Vyfiltrovat druhé kolo
  - Data – filtr – kolo – 2 – ctrl+a – ctrl+c – ctrl+n – ctrl+v
- Souhrn dle obce
  - Data – souhrn – u změny ve sloupci OBEC; použít funkci součet; označit všechny položky od VOL\_SEZNAM, kromě KC – ok
- Výsledek do nového sešitu
  - ctrl+a – ctrl+c – ctrl+n – klepnout pravým do první buňky a vložit jako hodnoty
- Upravit nový sešit
  - Vložit sloupec za sloupec OBEC – nový sloupec pojmenovat „Celkem“ – rozdělit sloupec obec (data – text do sloupců – oddělovač – mezera – ok) – seřadit dle celkem – smazat data v řádcích, které neobsahují slovo celkem - smazat prázdné sloupce, sloupce kde jsou jen nuly a sloupec celkem
  - Přejmenujte sloupce HLASY01, ...,HLASYX podle jmen uvedených v registrech
- ctrl+s – název: obvodN\_II
  - Před uložením přejmenovat také list
  - Název sešitu krátký, bez mezer a diakritiky

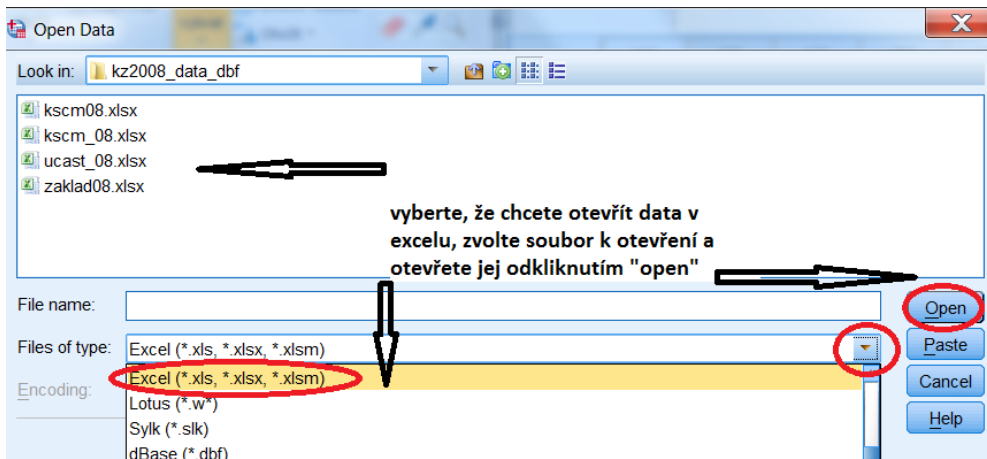
- V původním sešitu označit a smazat data za 2. kolo
- Zrušit filtr za kolo
- Upravit data za 1.kolo
- Souhrn dle obce
  - Data – souhrn – u změny ve sloupci OBEC; použít funkci součet; označit všechny položky od VOL\_SEZNAM, kromě KC – ok
- Výsledek do nového sešitu
  - ctrl+a – ctrl+c – ctrl+n – klepnout pravým do první buňky a vložit jako hodnoty
- Upravit nový sešit
  - Vložit sloupec za sloupec OBEC – nový sloupec pojmenovat „Celkem“ – rozdělit sloupec obec (data – text do sloupců – oddělovač – mezera – ok) – seřadit dle celkem smazat prázdné sloupce, sloupce kde jsou jen nuly a sloupec celkem
  - Přejmenujte sloupce HLASY01, ...,HLASYX podle jmen uvedených v registrech
  - Za název v každém sloupci přidejte „\_2“
- ctrl+s – název: obvodN\_II
  - Před uložením přejmenovat také list

# Spojení 1. a 2. kola a přidání dalších dat

- SPSS
- Přidání dat
  - Z předchozích voleb
  - Ze sčítání lidu
  - Z jiných zdrojů (MPSV, MŠMT, cokoli,...)



- Otevřete soubory, které chcete propojit, v SPSS



- Po otevření tabulek začněte se spojováním.

- Nejprve je nutné seřadit data v každé spojované tabulce podle sloupce OBEC

\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

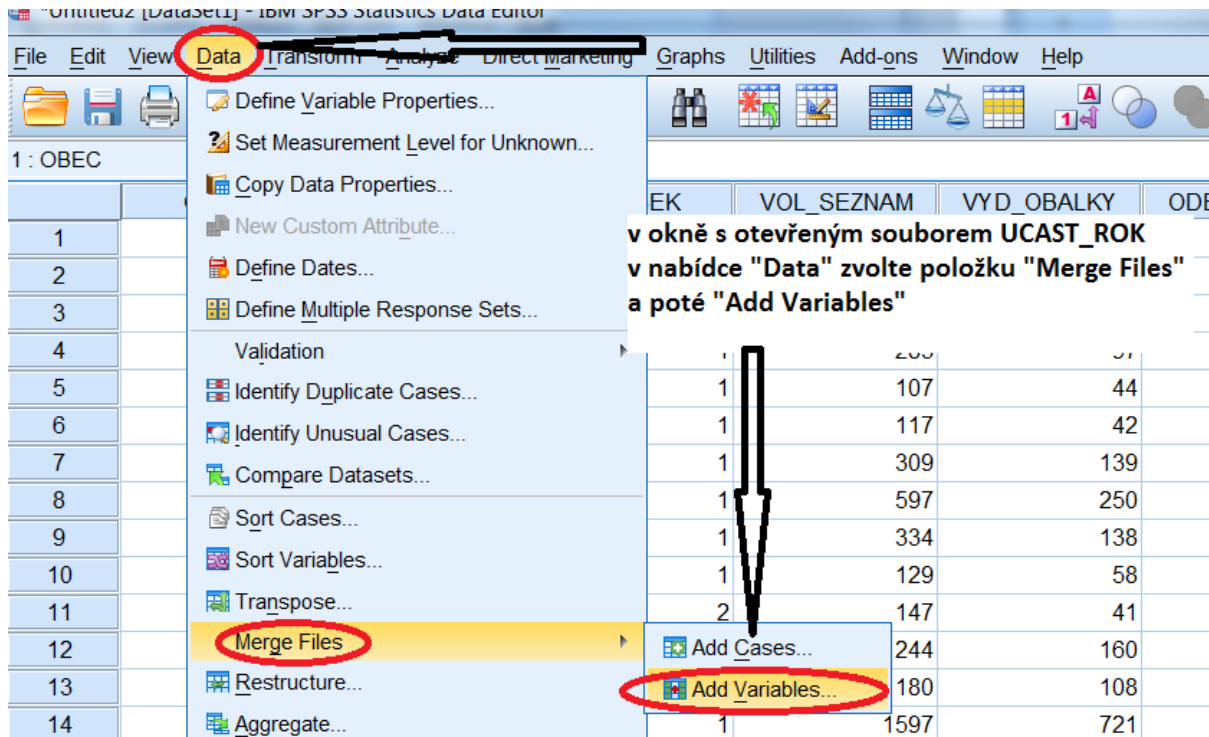
1 : OBEC 505927

	OBEC			BALKY	ODEVZ_OBAL	F
1	5059			15770	15526	
2	5061		1343	486	484	
3	5062		3564	1291	1277	
4	5064		1100	370	369	
5	5067		3497	1046	1041	
6	5067		2754	1141	1128	
7	5071		859	291	291	
8	5072		992	462	456	
9	5072		4389	1602	1595	
10	5073		1048	272	268	
11	507504	1	2664	765	753	
12	507580	1	5545	1480	1454	

Kliknout pravým a zvolit "sort ascending"

# Spojení dat

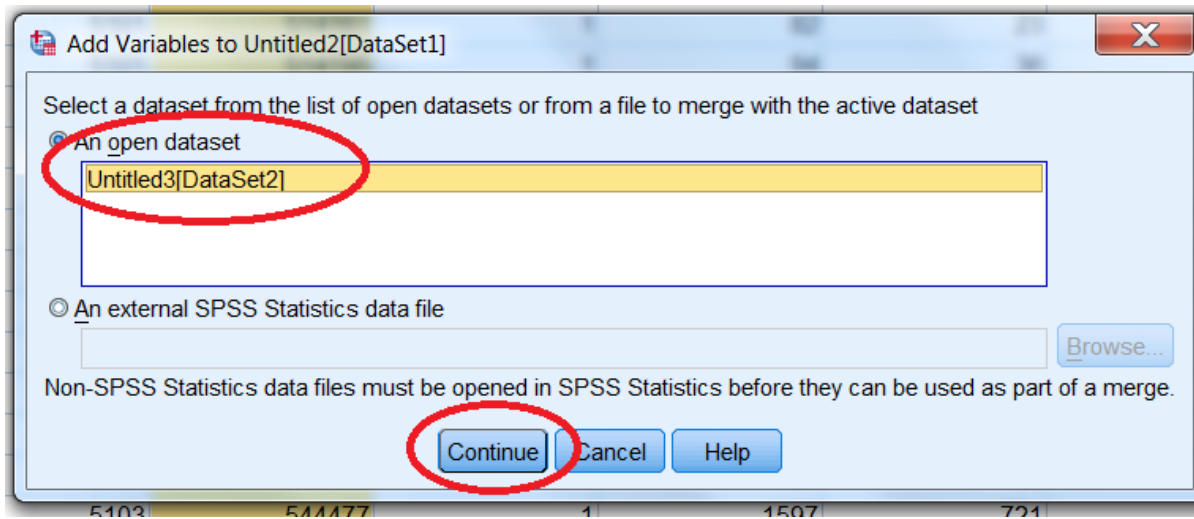
- Spojování začnete od 1. kola posledních voleb (v úkolu nejlépe od účasti v krajských volbách)



The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Data' menu is open, and the 'Merge Files' option is highlighted. The 'Add Variables...' option is also highlighted. A text box on the right provides instructions: 'v okně s otevřeným souborem UCAST\_ROK v nabídce "Data" zvolte položku "Merge Files" a poté "Add Variables"'. A large black arrow points from the 'Merge Files' option to the 'Add Variables...' option.

EK	VOL_SEZNAM	VYD_OBALKY	ODE
1	107	44	
1	117	42	
1	309	139	
1	597	250	
1	334	138	
1	129	58	
2	147	41	
	244	160	
	180	108	
1	1597	721	

- Vyberte



**Add Variables from DataSet2**

Excluded Variables:  
KOLO(\*)

New Active Dataset:  
V OBEC  
VYD\_OBALKY(\*)  
ODEVZ\_OBAL(\*)  
PL\_HL\_CELK(\*)  
Papuga(\*)  
Stiborský(\*)  
Skařupa(\*)  
Vlček(\*)  
Opálka(\*)

Key Variables:  
OBEC

**proměnnou OBEC  
přesuňte přes toto  
okno do pole Key  
Variables**

**Zaškrkněte obě pole**

Match cases on key variables  
 Cases are sorted in order of key variables in both datasets

Both files provide cases  
 Non-active dataset is keyed table  
 Active dataset is keyed table

Indicate case source as variable: source01

(\*)=Active dataset  
(+)=DataSet2

OK Paste Reset Cancel Help

509736

1

1196

486

481

415



# Sčítání lidu

- [https://www.czso.cz/csu/czso/otevrena\\_data\\_pro\\_vysl\\_edky\\_scitani\\_lidu\\_domu\\_a\\_bytu\\_2011\\_slodb\\_2011-](https://www.czso.cz/csu/czso/otevrena_data_pro_vysl_edky_scitani_lidu_domu_a_bytu_2011_slodb_2011-)
- **Výběr údajů ze SLDB 2011 za obyvatelstvo, domy a byty, domácnosti a vyjížďku**
  - Popis dat
  - Obyvatelstvo (excel v csv. složce)
    - Po otevření: data – text do sloupců – oddělovač: čárka
    - Vyfiltrování obcí: data – filtr – typuz\_naz: obec
    - Vykopírování do nového sešitu
    - Přejmenování sloupců (pomocí souboru popis dat)
    - Zjednodušení a zkrácení názvů! (jinak problémy v SPSS)
    - Smazání údajů jen za ženy/muže
    - **Výběr sloupců závisí na teoretických předpokladech**

- Spojení stejným způsobem jako výše
- Export do isu
- Smazání obcí „navíc“
- Vypočtení procent

# Nezaměstnanost

- <https://portal.mpsv.cz/sz/stat/nz/uzem>
- Data neobsahují identifikační kód
- Stažení dat za každý okres, do kterého zasahuje senátní obvod
- Připsání identifikace okresu do sloupce tabulky
- Seřazení dle okresu a obce
- Připsání kódu ze struktury území  
([https://www.czso.cz/csu/czso/i\\_zakladni\\_uzemni\\_i\\_ciselniky\\_na\\_uzemi\\_cr\\_a\\_klasifikace\\_cz\\_nuts](https://www.czso.cz/csu/czso/i_zakladni_uzemni_i_ciselniky_na_uzemi_cr_a_klasifikace_cz_nuts))

# Úkol

- Vytvořit datovou matici obsahující:
  - Podporu kandidátů v senátních volbách 2006 a 2012
  - Podporu stran ve volbách do Poslanecké sněmovny v roce 2013 a 2017 (pozor na názvy! – za každý název umístěte něco jako „\_PS13“)
  - Podporu stran v krajských volbách 2012 a 2016
  - Vybraná data ze sčítání lidu a jiných zdrojů
- A4 vysvětlující, proč jste zvolili dané proměnné (= teoretické předpoklady/odkazy k teorii)
- Odevzdání do půlnoci 1. 4. 2018.

# Četba k příštímu semináři

- Parker – Asencio: Gis and spatial analysis for social sciences: coding, mapping and modeling
  - Xiii – xvi: Overview
  - 1 – 24: o GISu
  - **51 – 83: Thematic maps**