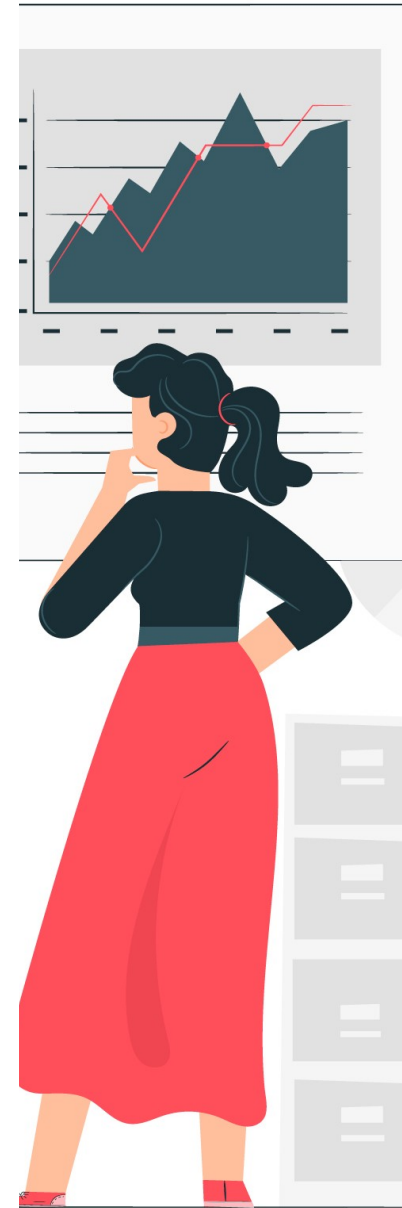




Logika kvantitativního výzkumu

POLb1006 Jak zkoumat politiku

Jakub Jusko



Podzim 2023

KVANTI VÝZKUM

- **Umožňuje výzkumníkům systematicky analyzovat komplexní politické fenomény prostřednictvím data-driven metod**
- Pracuje s fakty a ověřitelnými informacemi
- Dá se verifikovat (a často i replikovat)
- Výsledky se dají použít v praxi
- Doplnuje kvali výzkum

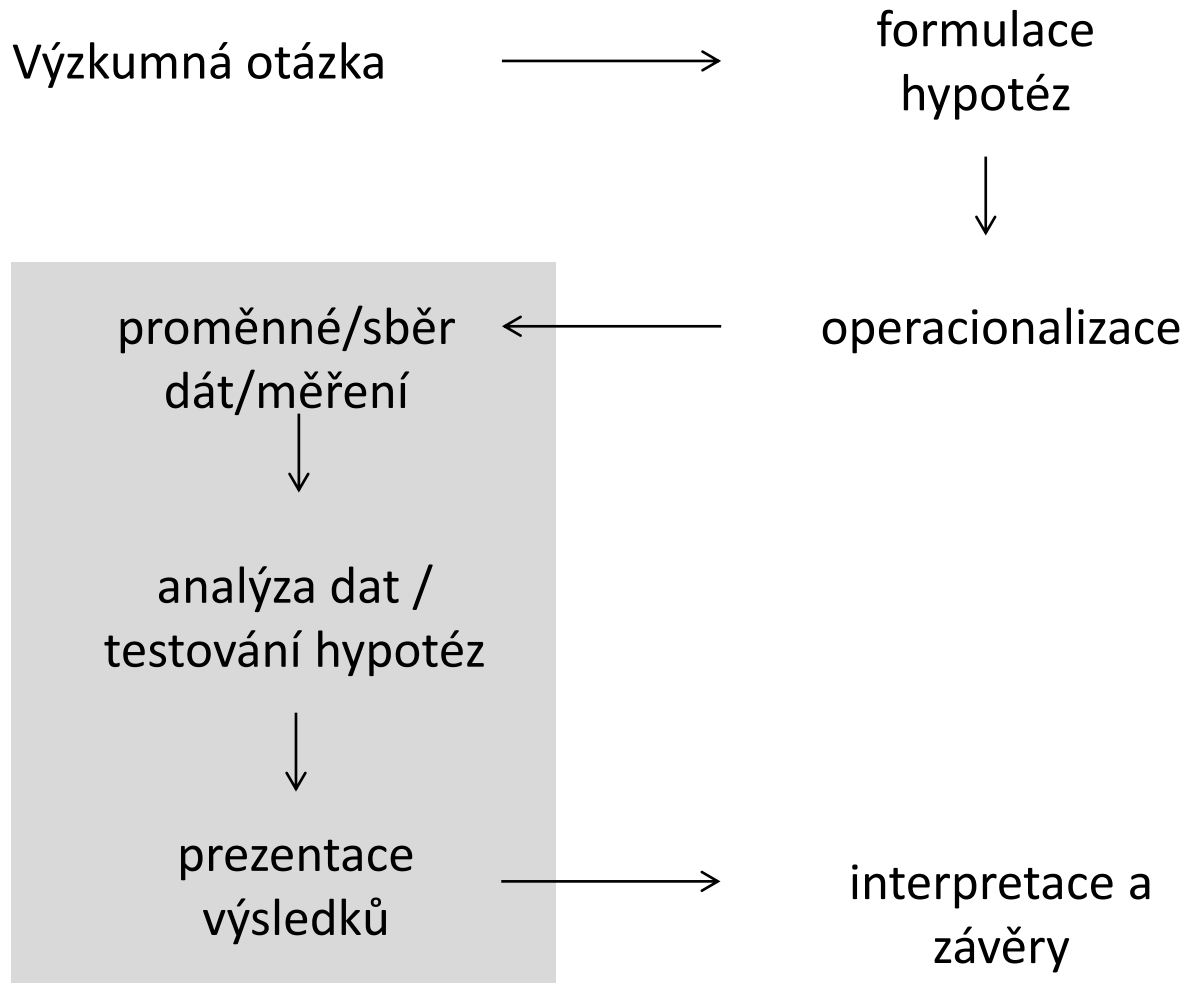
ALE někdy těžké na pochopení, možná subjektivita, \$

KVANTI VÝZKUM

! metoda, zahrnuje systematický **sběr** a **analýzu** číselných údajů -> ověření **hypotéz** + zodpovězení **VO** (zdůrazňuje objektivní a empirickou povahu)

*Rola **hypotéz** při vedení kvantitativního výzkumu, výzkumní pracovníci formulují kvalifikované odhady vztahů mezi proměnnými, které ověřují pomocí dat.

VÝZKUMNÝ PROCES



STATISTIKA JAKO NÁSTROJ ...

... na zodpovězení výzkumné otázky

Sama o sobě nestačí!

- kvalita výstupů závisí na kvalitě vstupů
- kvalitní výzkumný dizajn, korektní měření (kvalitní data), ...
- **správné pochopení výsledků**, jejich interpretace a reportování

Co však můžeme získat?

- překonání problémů s kauzalitou
(„kauzální překážky“)
- na základě malého vzorku usuzovat o celé populaci

DATA



NEŽ ZAČNEME ANALYZOVAT...

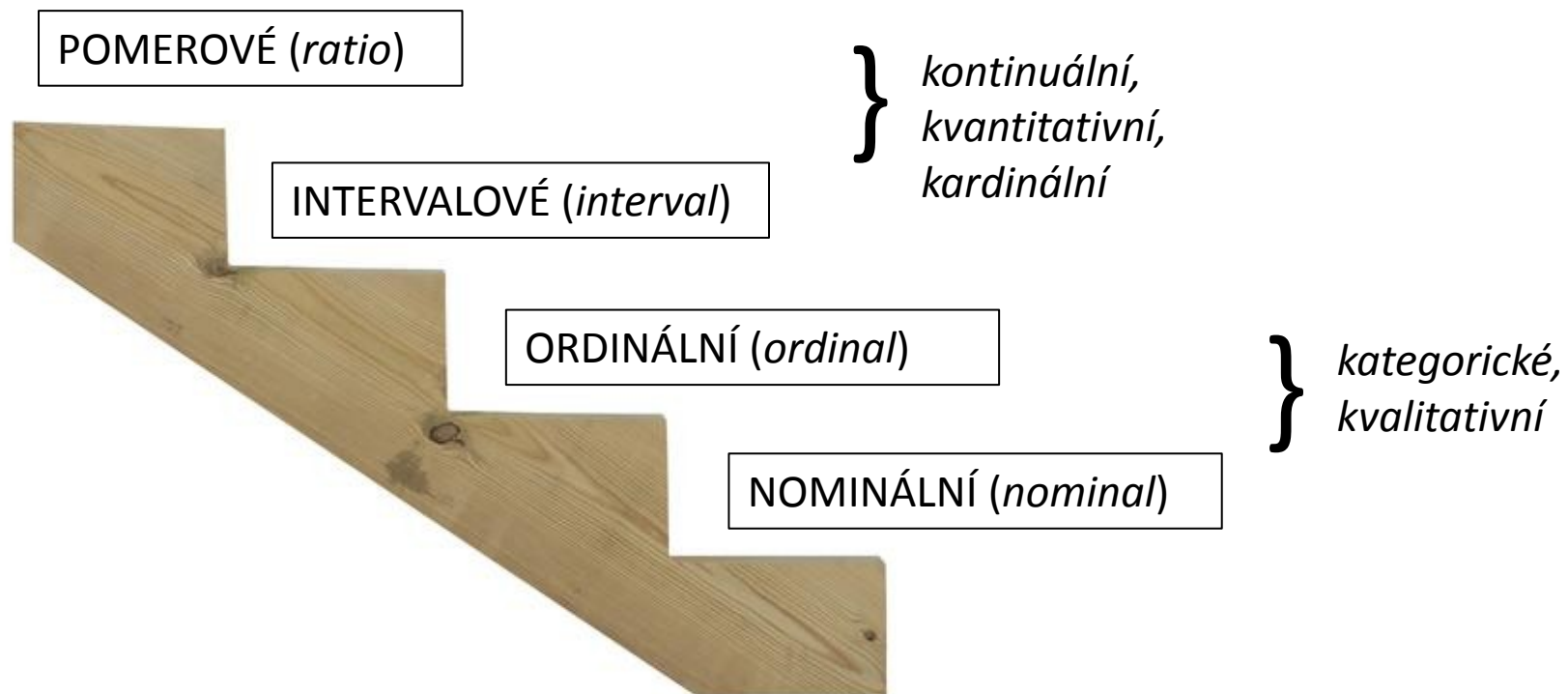
- potřebujeme **data** - **nevyhnutné** pro testování hypotéz
- OTÁZKA: 1) co měřit?
 - » závislá vs. nezávislá proměnná (?)
- 2) **jak měřit?**

ÚROVEŇ MĚŘENÍ

- „level of measurement“
- úroveň měření :
 - určuje povahu a množství informací, které proměnná obsahuje
 - ovlivňuje způsob, jakým data analyzujeme a interpretujeme
 - definuje typ proměnné

ÚROVEŇ MĚŘENÍ | Typy proměnných

- základné rozdění:



ÚROVEŇ MĚŘENÍ | Nominální proměnné

- **kvalitativní** odlišení (klasifikace) naměřených hodnot – označení vzájemně **oddělených kategorií, názvy objektů**
- naměřené hodnoty můžeme zařadit do určité **NEnumerické** kategorie, ale nemůžeme je mezi sebou kvantitativně porovnávat /řadit
- např. národnost, kraj v ČR, příslušnost poslance k pol. straně, pohlaví, barva očí, ...
- nepracujeme s nimi kvantitativním způsobem, **neumožňují** matematické operace – **ALE!**
- příklad z exit pollu (?)

ÚROVEŇ MĚŘENÍ | Ordinální proměnné

- stejně jako nominal. označují **kategorie**
- rozdíl oproti nominal.: **hodnoty mohou být seřazené** (podle určité charakteristiky)
- **nevíme** však „**o kolik**“ nebo „**kolik krát**“ je daná hodnota větší/menší jako jiná
- Např. míra spokojenosti (1. velmi spokojený, 2. spokojený, 3. nespokojený), závěrečné hodnocení kurzu (A, B, C, D, E, F), ...
- **příklad z exit pollu (?)**

ÚROVEŇ MĚŘENÍ | Intervalové proměnné

- **Umíme seřadit**, na rozdíl od ordinální **můžeme** u intervalové proměnné vypočítat i „**o kolik**“ je jedna hodnota větší/menší jako druhá
- hodnotami jsou **čísla**, při kterých poznáme jednotku měření, a vzdálenost mezi jednotlivými možnými hodnotami měření (použité škály) je **stejná**
- **nemají přirozenou absolutní nulu** (nula neznamena absenci měřené veličiny, je to jen arbitrárně učený bod na škále)
- Např. rok narození (letopočet), teplota Celzia, hodiny

- příklad z exit pollu (?)

ÚROVEŇ MĚŘENÍ | Poměrové proměnné

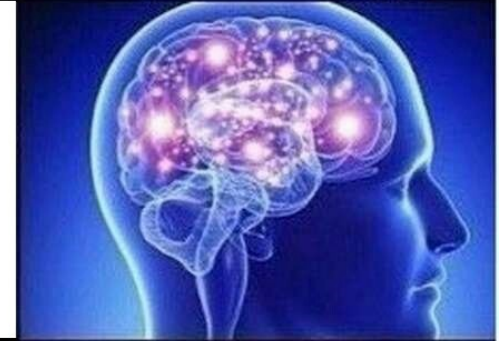
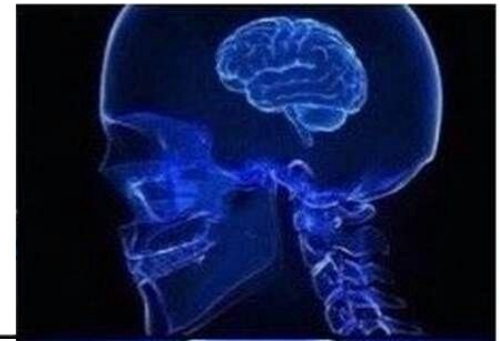
- stejné vlastnosti jako intervalová + **absolutní nula**
- „**nula**“ – skutečná absence měřené vlastnosti
- nejvyšší informační hodnota
- všechny matematické operace ($=$, \neq , $<$, $>$, $+$, $-$, $*$, $/$)
- např. počet bodů ze zkoušky, finance vynaložené na předvolební kampaň ...
- příklad z exit pollu (?)

NOMINÁLNÍ

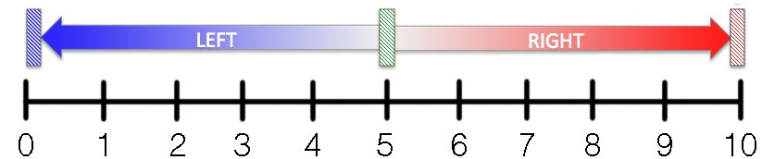
ORDINÁLNÍ

INTERVALOVÉ

POMĚROVÉ



ÚROVEŇ MĚŘENÍ | Test



ÚROVEŇ MĚŘENÍ | Test



ordinální

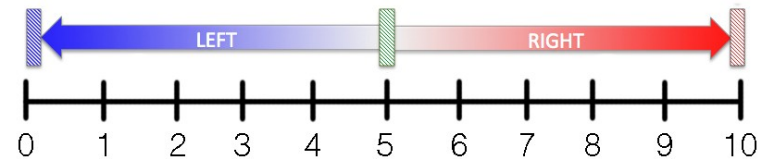


nominální



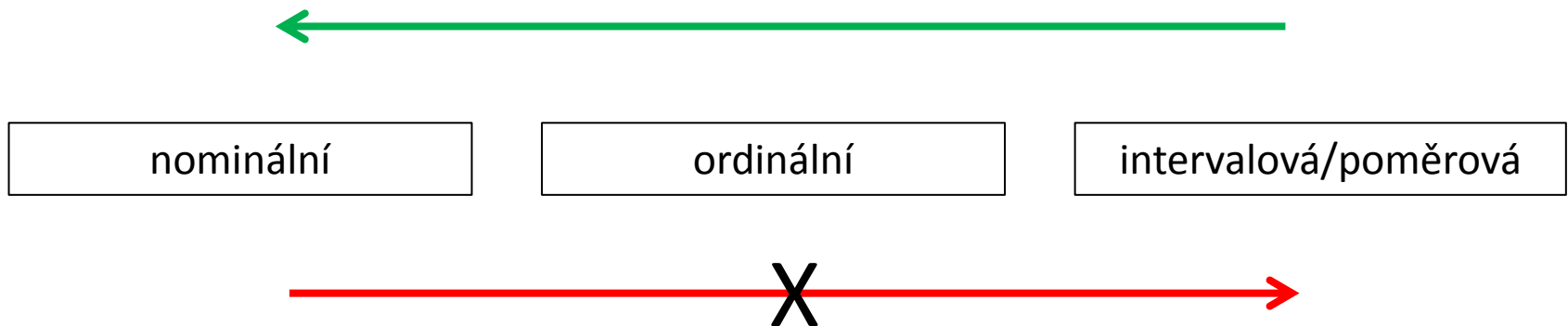
poměrová

intervalová



ÚROVEŇ MĚŘENÍ | Proč je to důležité?

- **důležité rozlišovat** mezi úrovněmi měření/typy proměnných -> různé typy proměnných -> různé statistické operace
- čím vyšší úroveň, tím lepší
- konverze



ÚROVEŇ MĚŘENÍ | Konverze

věk v letech:

16, 20, 31, 49, 52

typ proměnné ?

POMĚROVÁ

ORDINÁLNÍ

NOMINÁLNÍ

?

?

ÚROVEŇ MĚŘENÍ | Konverze

věk v letech:

16, 20, 31, 49, 52

typ proměnné ?

POMĚROVÁ

ORDINÁLNÍ

NOMINÁLNÍ

?

?

<20, 20-30, 31-40, 41-50, >50

- adolescenti
- dospívající
- mladí



Docházka

DATOVÁ MATICE

sloupce = hodnoty
proměnné

	Q1	Q3	Q4	Q4f	Q5	Q6	Q6J
1	spíše ne	99	neví	nedokáž	trvalé byd...	neví	x
2	spíše ne	6	před více než 2 týdny	99	trvalé byd...	Fiala	x
4	spíše ne	pr...	dlouhodobě rozhodn...	99	trvalé byd...	někdo jiný	Robejšek
5	rozhodně ano	4	v posledních 2 dnech	99	trvalé byd...	Fiala	x
6	spíše ano	99	dlouhodobě rozhodn...	99	trvalé byd...	Bartoš	x
7	ani ano, ani ne	pr...	před více než 3 měsíci	99	trvalé byd...	neví	x
8	spíše ano	5	dlouhodobě rozhodn...	99	trvalé byd...	Babiš	X
9	rozhodně ne	5	před více než 3 měsíci	99	trvalé byd...	neví	x
10	ani ano, ani ne	99	v posledních 2 dnech	99	trvalé byd...	Zaorálek	x
11	rozhodně ano	7	před více než 2 týdny	99	trvalé byd...	Fiala	x
12	rozhodně ano	8	před více než 3 měsíci	99	trvalé byd...	neví	x
13	rozhodně ano	8	dlouhodobě rozhodn...	99	trvalé byd...	Fiala	x
14	rozhodně ano	8	dlouhodobě rozhodn...	99	trvalé byd...	Kalousek	x
15	spíše ano	8	v posledních 2 týdne...	99	trvalé byd...	někdo jiný	Klaus
16	rozhodně ano	7	v posledních 2 dnech	99	trvalé byd...	neví	X
17	ani ano, ani ne	5	před více než 2 týdny	99	trvalé byd...	Bartoš	x
18	spíše ano	99	dlouhodobě rozhodn...	99	neví	neví	x
19	ani ano, ani ne	le...	před více než 3 měsíci	99	trvalé byd...	neví	x
20	spíše ano	6	v posledních 2 týdne...	99	trvalé byd...	neví	x
21	spíše ne	7	před více než 3 měsíci	99	trvalé byd...	neví	x
22	ani ano, ani ne	8	v posledních 2 dnech	99	trvalé byd...	Babiš	x

řádky = jednotlivé
případy

DATOVÁ MATICE

sloupce = hodnoty
proměnné

	Q1	Q3	Q4	Q4f	Q5	Q6	Q6J
1	4	99	99	nedokáž	1	99	x
2	4	6	3	99	1	4	x
3	4	10	1	99	1	11	Robejšek
4	1	4	5	99	1	4	x
5	2	99	1	99	1	9	x
6	3	10	2	99	1	99	x
7	2	5	1	99	1	2	X
8	5	5	2	99	1	99	x
9	3	99	5	99	1	1	x
10	1	7	3	99	1	4	x
11	8	8	2	99	1	99	x
12	1	1	99	1	1	4	x
13	2	8	4	99	1	11	Klaus
14	1	7	5	99	1	99	X
15	3	5	3	99	1	9	x
16	2	99	1	99	99	99	x
17	3	0	2	99	1	99	x
18	2	6	4	99	1	99	x
19	4	7	2	99	1	99	x
20	3	8	5	99	1	2	x
21							
22							

řádky = jednotlivé
případy

DATOVÁ MATICE vs. Tabulka

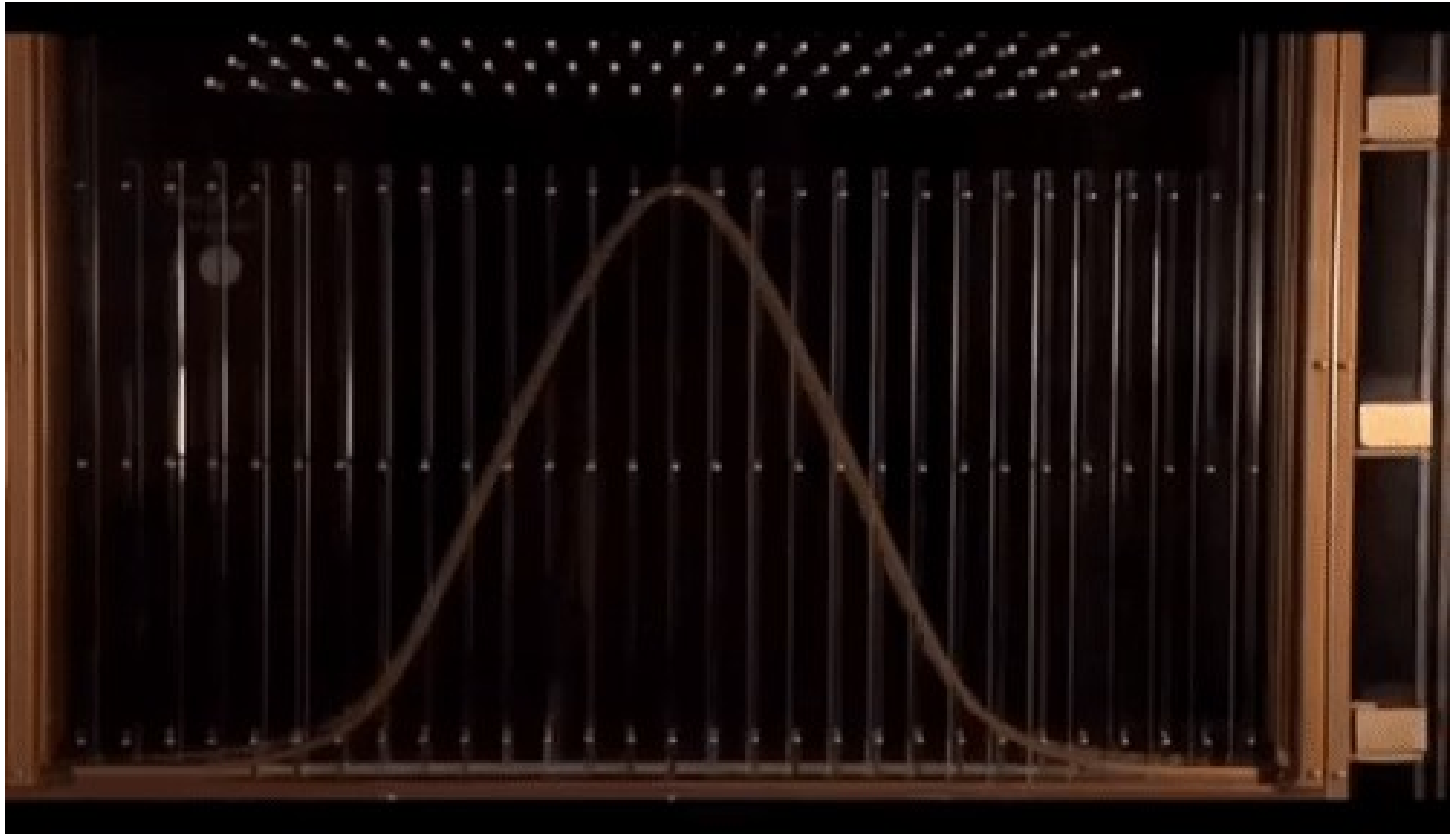
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Kraj	Okresy	Zapísaní v	Zúčastnení	Účasť	Odovzdané	Odovzdaná	Platné hlas	Platné hlas	Počet pos	Počet zvo	Invalidity	Invalidity	Rok
2	Banskobystrí	915	523 084	97 581	18.7	97 540	99.95	92 166	95 017	49	49	5.5	2.6	2005
3	Banskobystrí	913	511 735	123 683	24.2	123 637	99.96	116 806	119 608	49	49	5.5	3.3	2001
4	Banskobystrí	912	530 301	130 435	24.6	130 378	99.95	123 223	126 545	49	49	5.5	2.9	2013
5	Banskobystrí	918	530 098	143 445	27.1	143 397	99.96	135 826	139 344	49	49	5.3	2.8	2009
6	Banskobystrí	915	526 978	212 360	40.3	212 286	99.96	204 321	204 652	49	49	3.8	3.6	2017
7	Bratislavský	527	525 536	75 977	14.5	75 948	99.96	73 786	73 466	50	50	2.8	3.3	2005
8	Bratislavský	541	545 707	106 236	19.5	106 199	99.96	103 641	103 624	44	44	2.4	2.4	2009
9	Bratislavský	554	563 131	121 928	21.7	121 877	99.95	119 422	116 853	44	44	2.0	4.1	2013
10	Bratislavský	514	498 647	119 519	24.0	119 460	99.95	116 642	115 956	46	46	2.4	2.9	2001
11	Bratislavský	569	584 995	183 342	31.3	183 245	99.94	180 490	176 100	50	50	1.5	3.9	2017
12	Košický	830	601 029	115 877	19.3	115 825	99.95	107 984	112 034	57	57	6.8	3.3	2005
13	Košický	851	624 153	110 937	17.8	110 890	99.95	99 818	107 835	57	57	10.0	2.8	2013
14	Košický	826	581 798	126 803	21.8	126 754	99.96	117 337	121 685	57	57	7.4	4.0	2001
15	Košický	846	616 847	141 487	22.9	141 420	99.95	128 596	137 724	57	57	9.1	2.6	2009
16	Košický	854	630 178	168 454	26.7	168 392	99.96	157 651	163 327	57	57	6.4	3.0	2017
17	Nitriansky	743	572 536	158 467	27.7	158 416	99.96	154 919	154 729	52	52	2.2	2.3	2005
18	Nitriansky	743	584 770	104 706	17.9	104 688	99.98	100 543	102 053	54	54	4.0	2.5	2013
19	Nitriansky	737	582 148	127 019	21.8	126 964	99.95	123 172	124 270	54	54	3.0	2.1	2009
20	Nitriansky	747	582 228	156 268	26.8	156 213	99.96	152 993	151 795	54	54	2.1	2.8	2017
21	Nitriansky	749	559 940	194 267	34.7	194 194	99.96	186 687	188 248	52	52	3.9	3.1	2001
22	Prešovský	989	593 025	115 513	19.5	115 453	99.94	106 515	111 965	62	62	7.7	3.0	2005
23	Prešovský	1 003	630 452	139 525	22.1	139 422	99.92	129 145	135 819	62	62	7.4	2.6	2013
24	Prešovský	975	568 685	145 051	25.5	145 002	99.96	134 183	138 799	60	60	7.5	4.3	2001
25	Prešovský	995	614 500	161 736	26.3	161 680	99.96	151 413	156 929	62	62	6.4	2.9	2009
26	Prešovský	1 009	639 324	187 983	29.4	187 922	99.96	179 289	183 618	62	62	4.6	2.3	2017
27	Spolu za SR	5 852	4 282 070	771 951	18.0	771 635	99.95	732 547	749 919	412	412	5.1	2.8	2005
28	Spolu za SR	5 932	4 463 039	897 536	20.1	897 161	99.95	849 101	871 762	408	408	5.4	2.8	2013
29	Spolu za SR	5 902	4 397 619	1 007 056	22.9	1 006 692	99.96	955 649	981 241	408	408	5.1	2.5	2009
30	Spolu za SR	5 812	4 141 572	1 077 828	26.0	1 077 434	99.96	1 024 878	1 040 620	401	401	4.9	3.4	2001
31	Spolu za SR	5 968	4 500 142	1 348 114	29.9	1 347 550	99.95	1 301 445	1 308 462	416	416	3.4	2.9	2017
32	Trenčiansky	639	483 599	59 512	12.3	59 480	99.94	55 778	57 462	45	45	6.2	3.4	2005
33	Trenčiansky	645	496 864	86 337	17.4	86 308	99.96	81 829	84 015	45	45	5.2	2.7	2013
34	Trenčiansky	648	493 810	101 682	20.6	101 640	99.95	96 062	99 090	45	45	5.5	2.5	2009
35	Trenčiansky	629	471 297	101 573	21.6	101 545	99.97	96 571	97 746	45	45	4.9	3.7	2001

DATOVÁ MATICE vs. Tabulka

	A	B	C	D	E	F	G
1	Volebná účasť v prvom kole krajských volieb na Slovensku						
2	kraj	Účasť 2001	Účasť 2005	Účasť 2009	Účasť 2013	Účasť 2017	
3	Banskobystrický	24.16	18.65	27.06	24.59	40.29	
4	Bratislavský	23.96	14.45	19.46	21.65	31.34	
5	Košický	21.79	19.27	22.93	17.77	26.73	
6	Nitriansky	34.69	27.67	21.81	17.90	26.83	
7	Prešovský	25.50	19.47	26.31	22.13	29.40	
8	Trenčiansky	21.55	12.30	20.59	17.37	26.32	
9	Trnavský	33.73	14.50	20.46	17.46	24.74	
10	Žilinský	23.47	15.69	23.68	21.57	33.84	
11	Spolu za SR	26.02	18.02	22.90	20.11	29.94	
12							

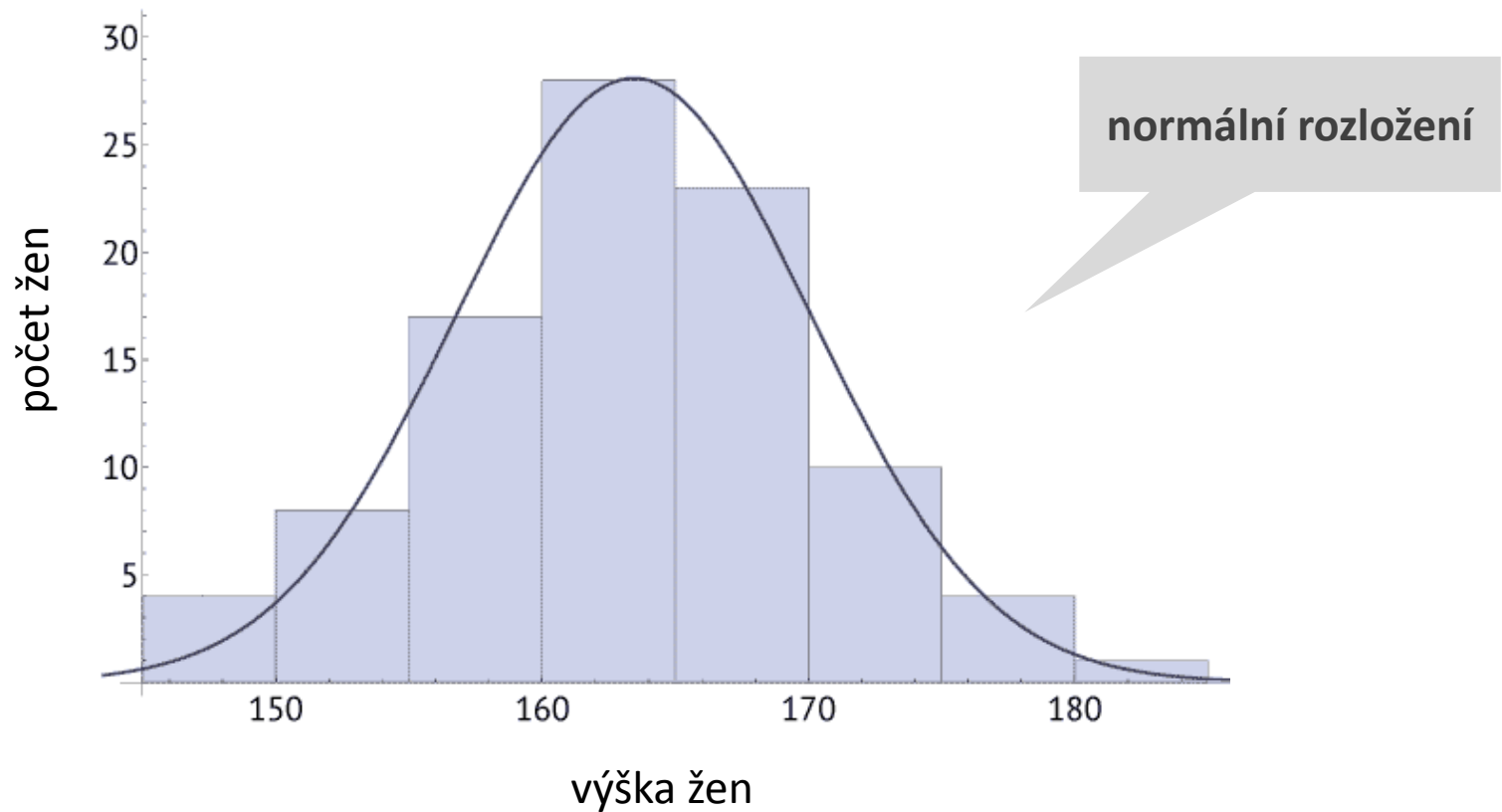
ANALÝZA DAT





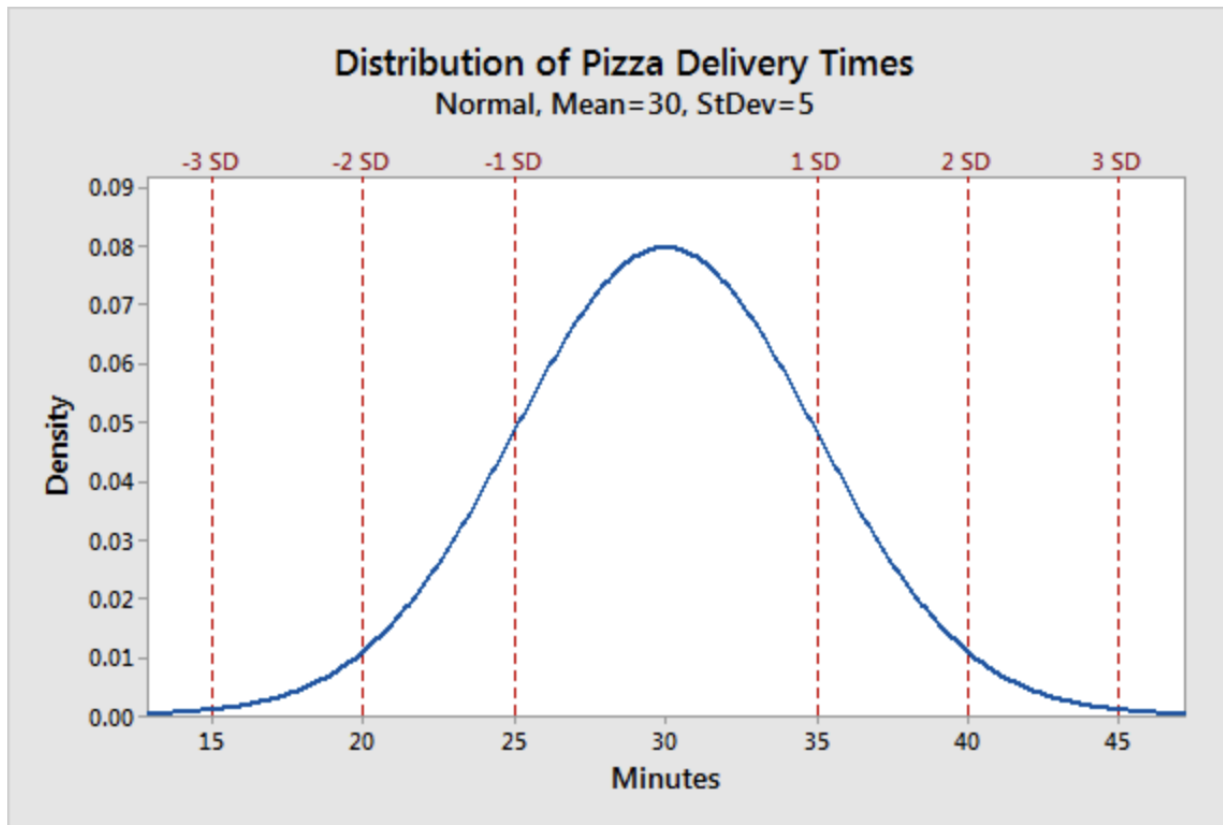
FREKVENCE VÝSKYTU

- grafické znázornění rozložení hodnot proměnné (trend)
- histogram



FREKVENCE VÝSKYTU

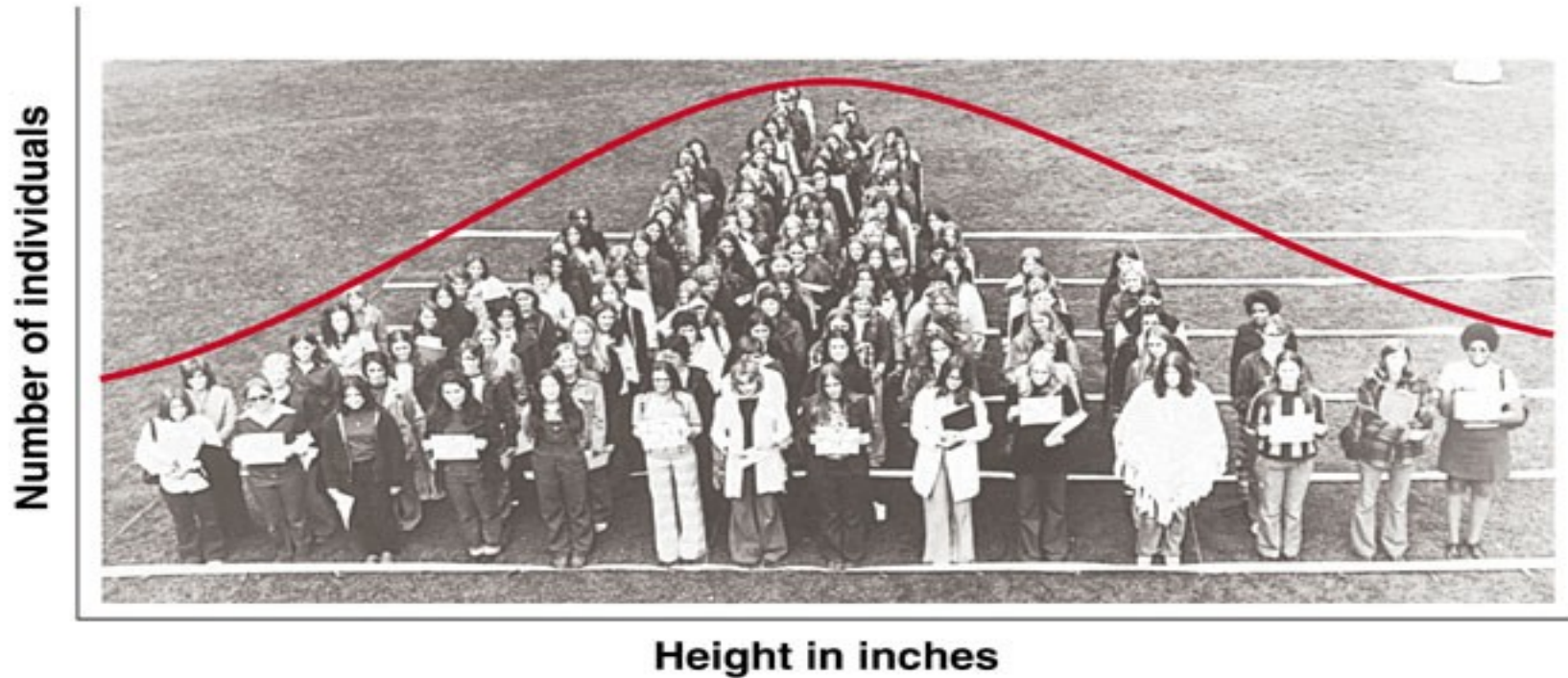
- grafické znázornění rozložení hodnot proměnné (trend)
- histogram



normální rozložení

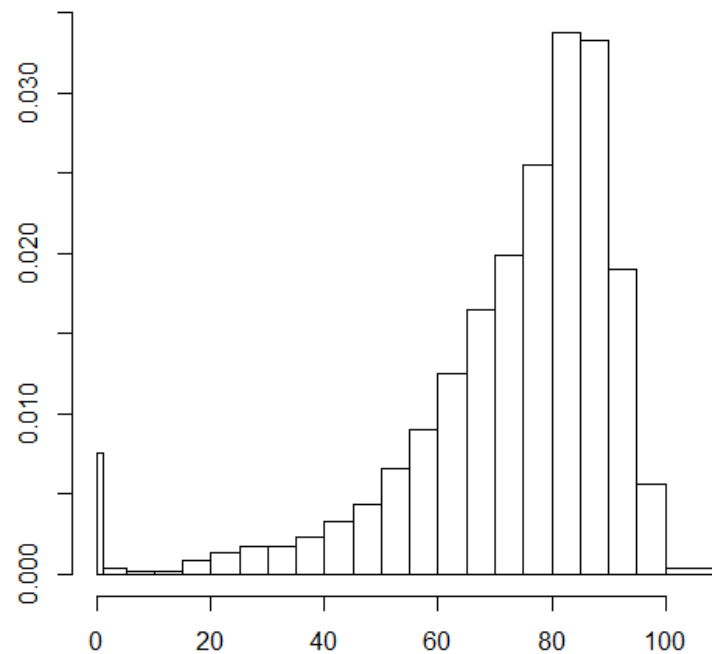
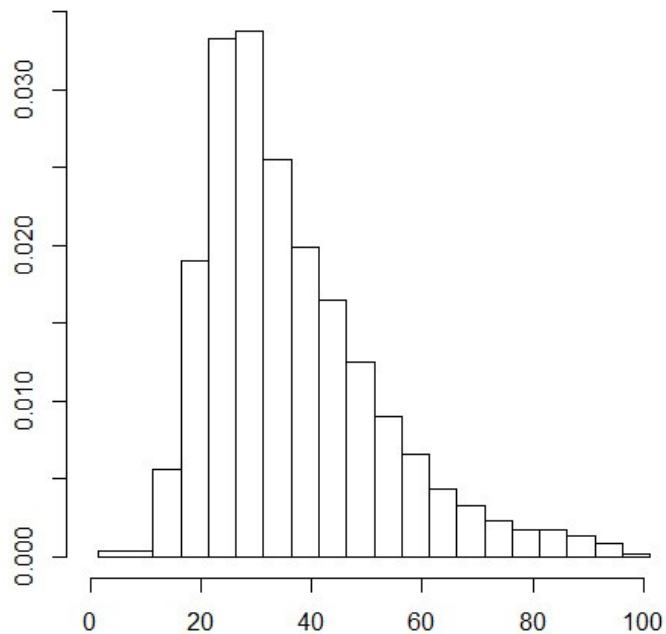
FREKVENCE VÝSKYTU

- grafické znázornění rozložení hodnot proměnné (trend)
- histogram



FREKVENCE VÝSKYTU

- zešikmené rozložení



STŘEDNÍ HODNOTY

- základní charakteristika dat, typická hodnota dané proměnné
- koncentrace informací do **jednoho čísla**
 - + **jednoduchý výpočet**
 - + **rychlá základní informace o rozložení dat proměnné**
 - **příliš zjednodušující**
 - **citlivost na extrémní hodnoty**

STŘEDNÍ HODNOTY

Modus

- nejčastější hodnota dané proměnné

13, 18, 13, 14, 13, 16, 14, 21, 13

Medián

- prostřední hodnota dané proměnné

13, 18, 13, 14, 13, 16, 14, 21, 13
13, 13, 13, 13, 14, 14, 16, 18, 21

Průměr

- součet hodnot vydělený počtem hodnot

NOMINÁLNÍ:

- modus

ORDINÁLNÍ:

- modus

- medián

INTERVALOVÉ:

- modus

- medián

- průměr

POMEROVÉ:

- modus

- medián

- průměr

STŘEDNÍ HODNOTY | Průměr

- velmi citlivý na extrémní hodnoty
- nemá smysl při asymetrickém rozložení dat
- může poskytovat informaci, která nedopovídá skutečnosti (!)

PRÍKLAD

- hodnocení úspěšnosti určité terapie
 - 20 pacientů přežilo měsíc, 1 pacient přežil 30 let (?)
 - Průměr dožití = 18 měsíců (1,5)

STŘEDNÍ HODNOTY | Průměr

- Průměr počtu vypitých piv za jeden týden

Jedinec	Muži (počet piv)	Ženy (počet piv)
Jedinec č. 1	8	0
Jedinec č. 2	8	0
Jedinec č. 3	8	0
Jedinec č. 4	8	0
Jedinec č. 5	8	40
Součet:	40	40
Aritmetický průměr:	8	8

STŘEDNÍ HODNOTY vs. FREKVENCE VÝSKYTU

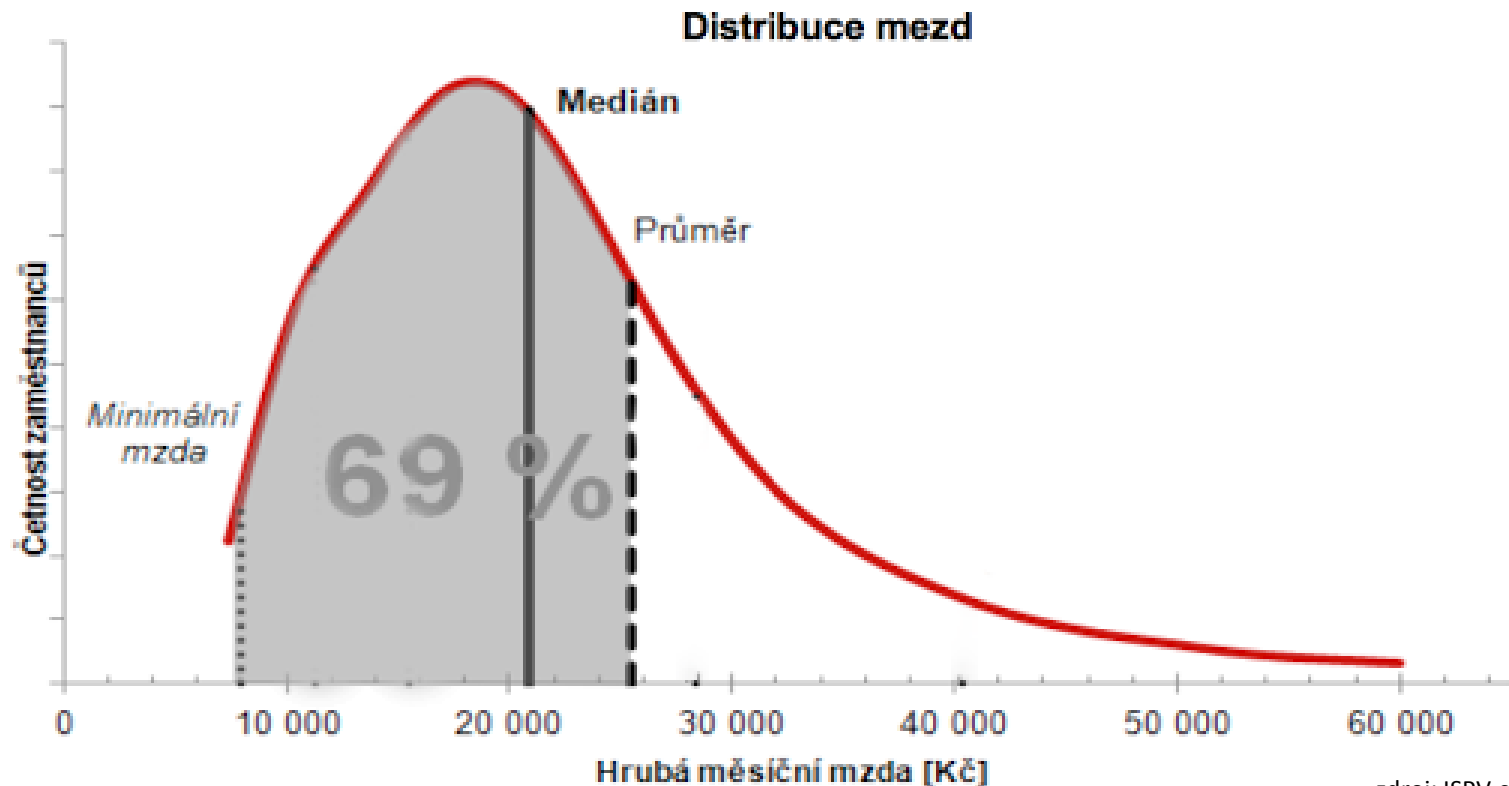
- Příjem za měsíc

jedinec č.1:	17 000,- Kč
jedinec č. 2:	18 000,- Kč
jedinec č. 3:	18 000,- Kč
jedinec č. 4:	19 000,- Kč
jedinec č. 5:	19 000,- Kč
jedinec č. 6:	19 000,- Kč
jedinec č. 7:	20 000,- Kč
jedinec č. 8:	20 000,- Kč
<u>jedinec č. 9:</u>	<u>21 000,- Kč</u>
Průměr:	19 000,- Kč
MODUS:	19 000,- Kč
MEDIÁN:	19 000,- Kč

jedinec č.1:	17 000,- Kč
jedinec č. 2:	18 000,- Kč
jedinec č. 3:	18 000,- Kč
jedinec č. 4:	19 000,- Kč
jedinec č. 5:	19 000,- Kč
jedinec č. 6:	19 000,- Kč
jedinec č. 7:	20 000,- Kč
jedinec č. 8:	20 000,- Kč
jedinec č. 9:	21 000,- Kč
jedinec č. 10:	50 000,- Kč
<u>jedinec č. 11:</u>	<u>100 000,- Kč</u>
Průměr:	29 200,- Kč
MODUS:	19 000,- Kč
MEDIÁN:	19 000,- Kč

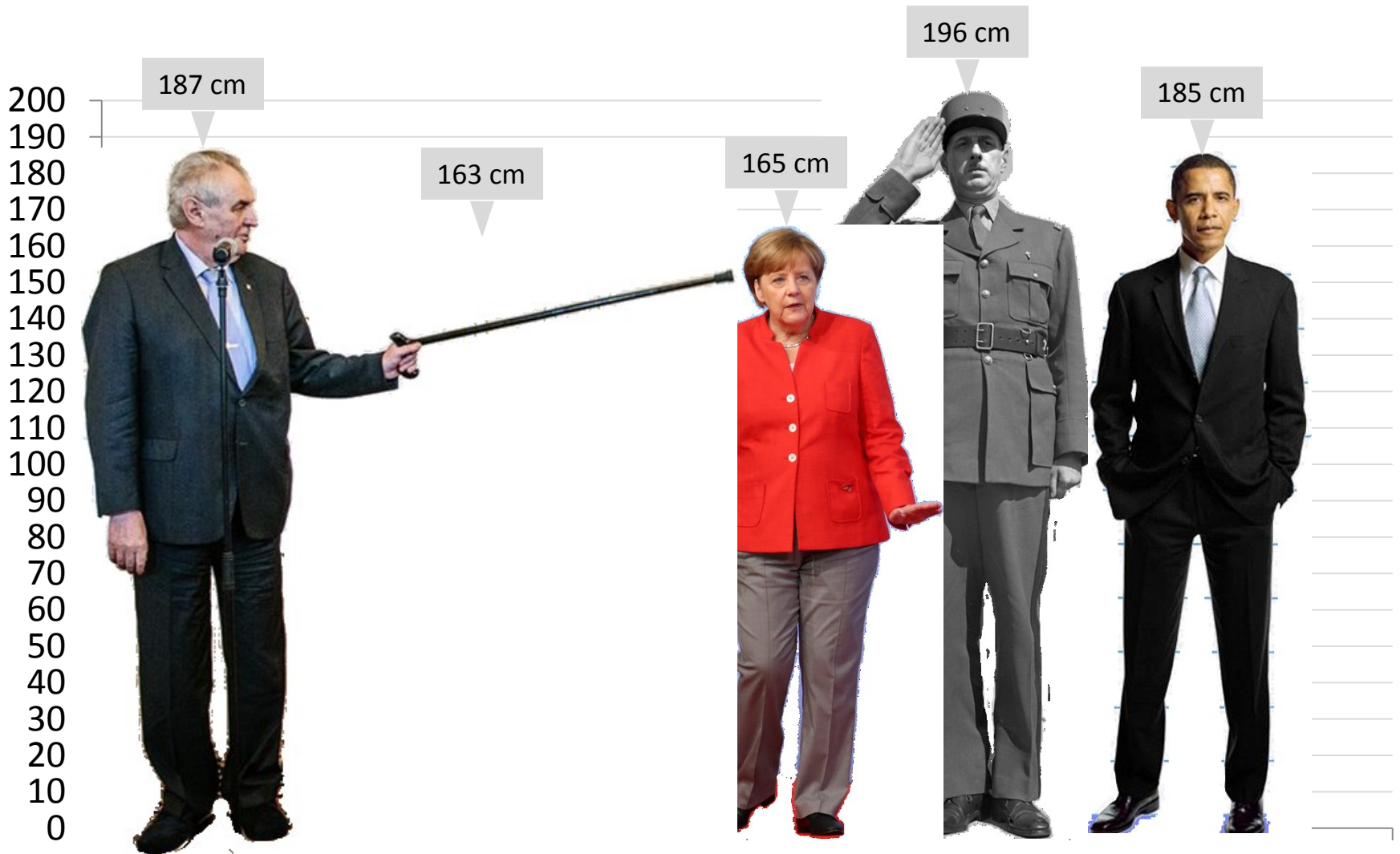
rozložení/histogram?

STŘEDNÍ HODNOTY vs. FREKVENCE VÝSKYTU



zdroj: ISPV.cz

- někdy vhodné využít statistiky nižší úrovně (modus, medián)
- **POZOR na homogenitu vzorku, VARIABILITU dat** (rozložení hodnot proměnné)

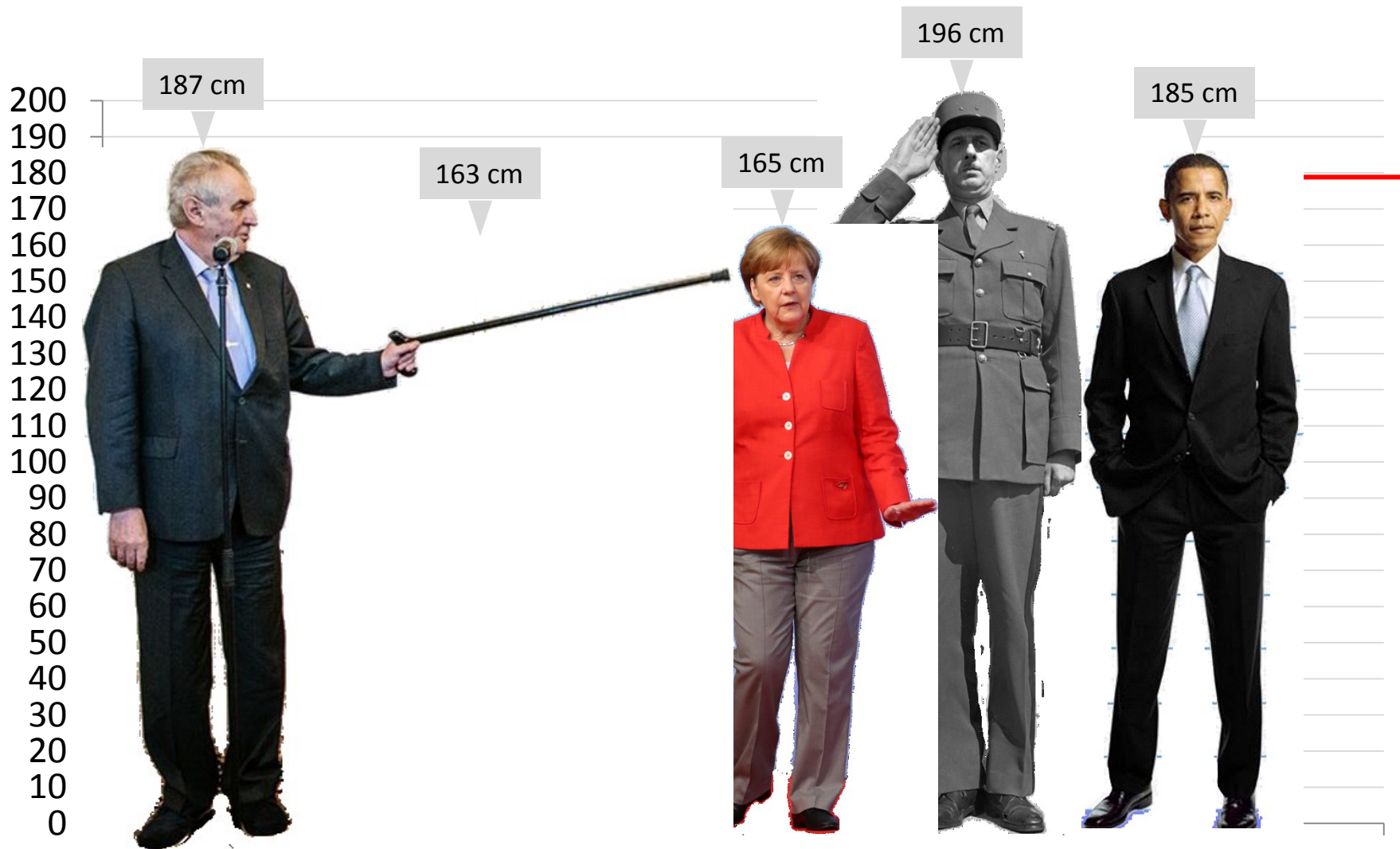


JAK MĚŘIT VARIABILITU?

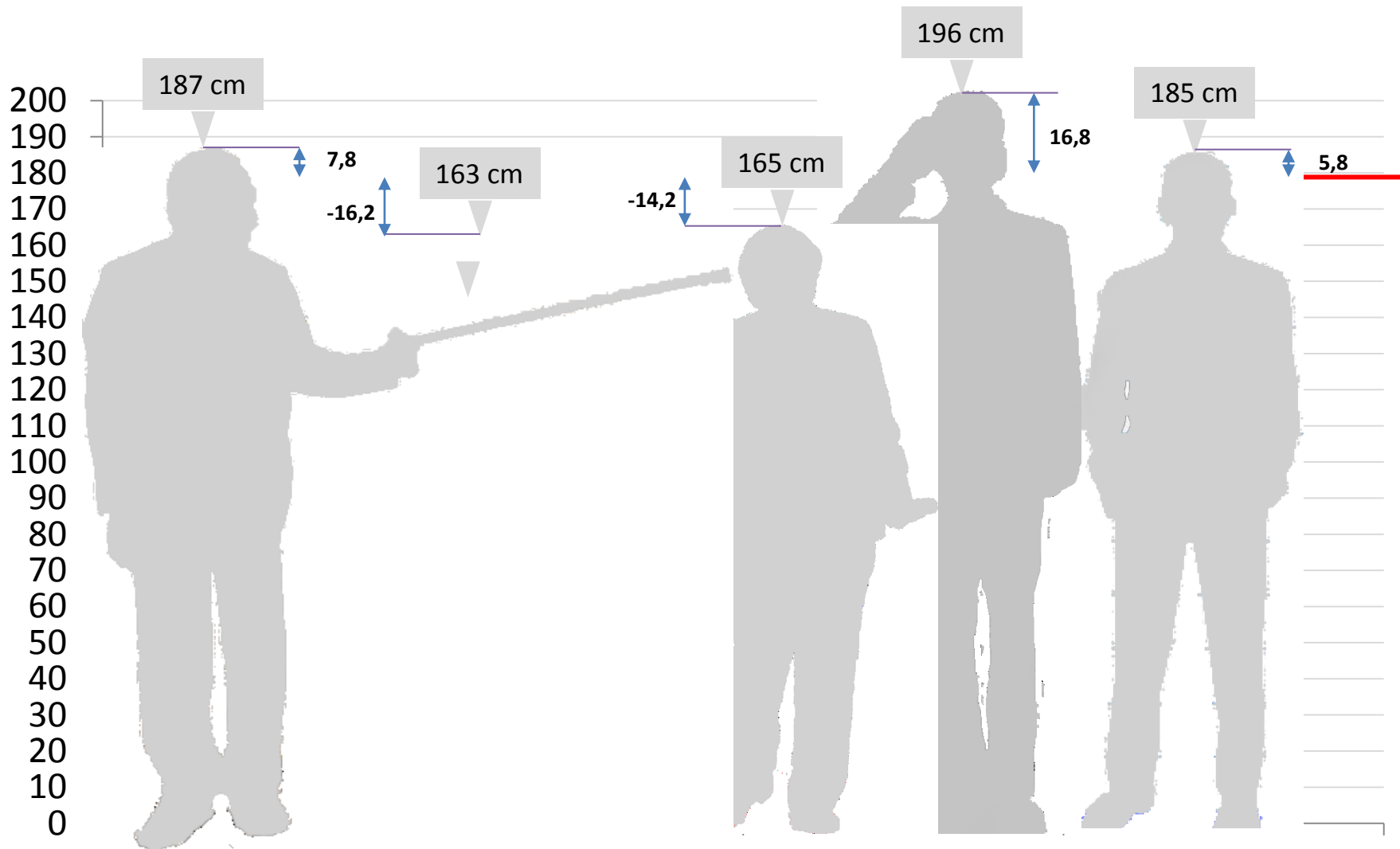
Rozptyl: průměr druhých mocnin vzdáleností hodnot od jejich průměru → **odchyly od průměru**

$$s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

Směrodatná odchylyka: $\sqrt{\text{rozptyl}}$



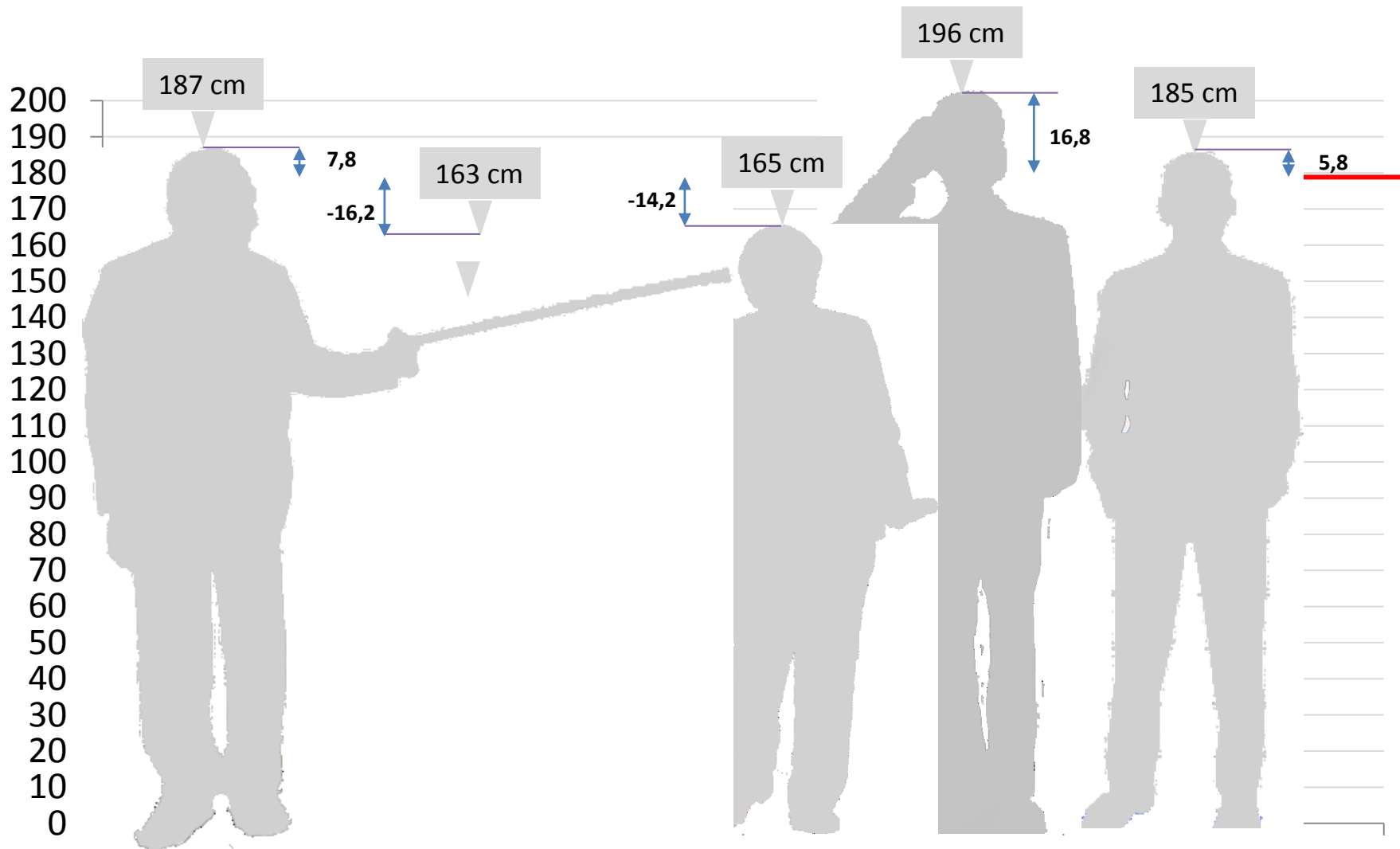
$$\begin{aligned}
 \text{průměr} &= \frac{187+163+165+196+185}{5} \\
 &= \frac{896}{5} \\
 &= 179,2
 \end{aligned}$$



$$s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

$$\text{rozptyl} = \frac{7,8^2 + (-16,2)^2 + (-14,2)^2 + 16,8^2 + 5,8^2}{5-1}$$

$$= 210,2$$



$$s^2 = \frac{\sum (X - \bar{X})^2}{N-1}$$

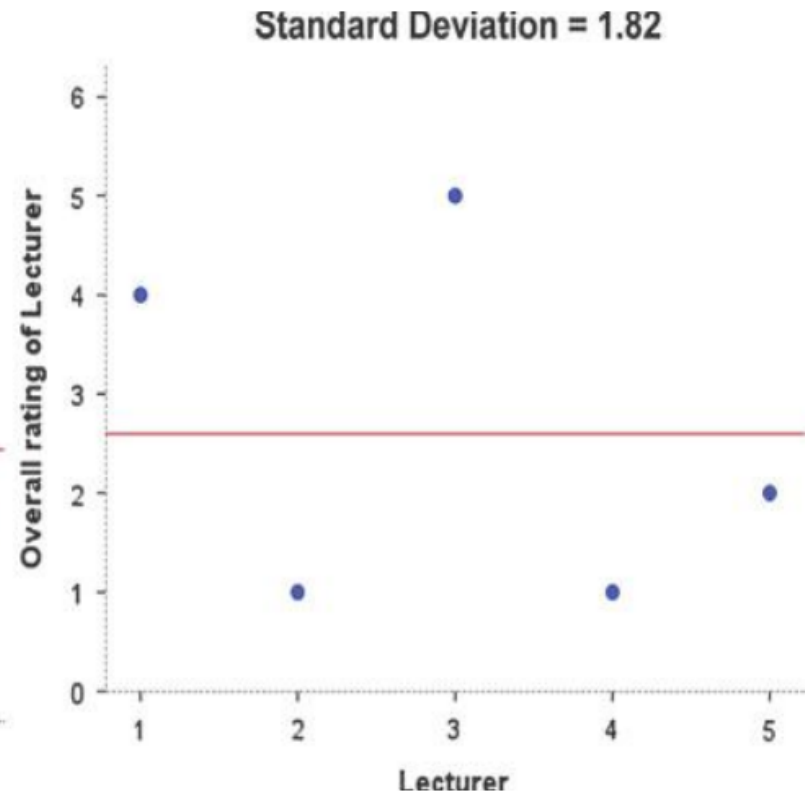
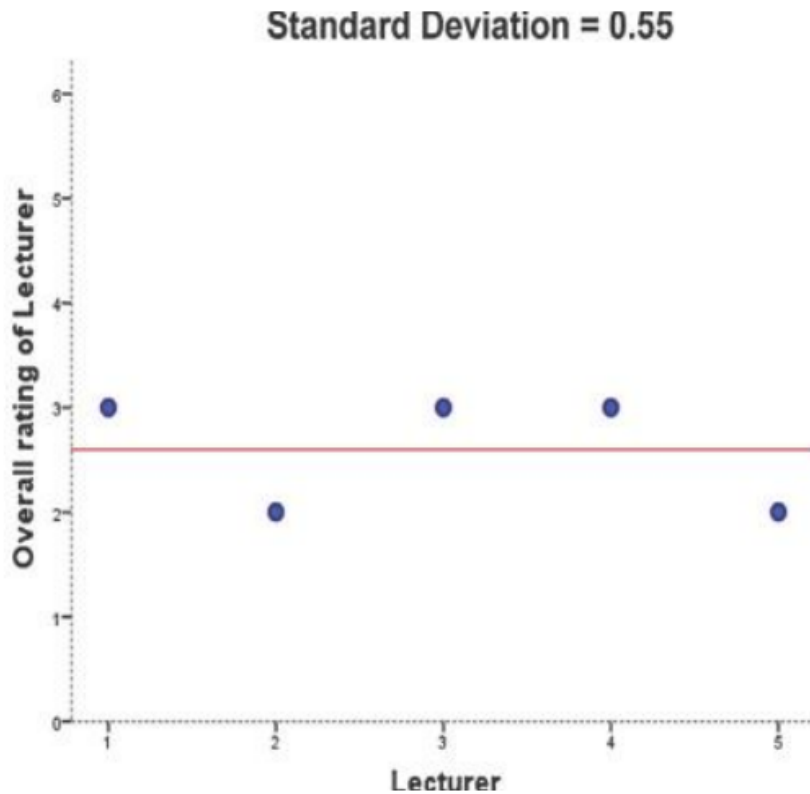
směrodatná odchylka

$$= \sqrt{210,2}$$

$$= 14,5$$

VARIABILITA

- Na co je to dobré?
- vzájemné porovnání dvou vzorek (směrodatná odchylka)



OD POPISOVÁNÍ K VYVOZOVÁNÍ

populace vs. vzorek

OD POPISOVÁNÍ K VYVOZOVÁNÍ

- DESKRIPTIVNÍ STATISTIKA
- **INFERENČNÍ STATISTIKA** – „statistics used to draw conclusions about significant relationships between variables“
- sociální vědy → zjištění založené na vzorkách → **snaha o generalizaci na celou populaci**
 - máme pod kontrolou všechny intervenující proměnné?
 - skutečně naše měření odpovídá konceptu?
- výsledky jsou doprovázeny **nejistotou**

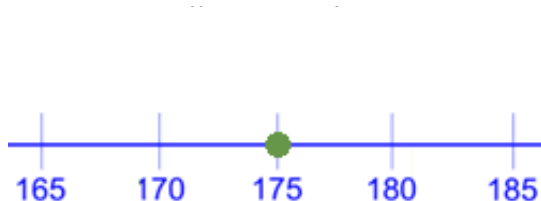
Cvičení

Cvičení

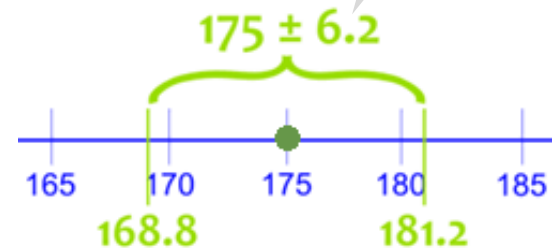
1. "Provádíte průzkum s cílem zjistit volební preference ve městě s 1 milionem obyvatel, ale máte omezený rozpočet na průzkum pouze 500 osob. Jak byste vybrali vzorek?"
2. "Chcete analyzovat názory studentů na FSS se třemi hlavními katedrami: politologie, psychologie a sociologie. Můžete provést průzkum u 100 studentů. Jak byste vybrali vzorek?"
3. "Zkoumáte politické postoje různých věkových skupin v jedné zemi. Vaše zdroje vám umožňují provést průzkum pouze u 50 respondentů. Jak byste vybrali svůj vzorek?"

OD POPISOVÁNÍ K VYVOZOVÁNÍ

- **interval spolehlivosti** – interval, ve kterém skutečná hodnota dané charakteristiky leží s určitou pravděpodobností (95%)
- změříme-li 100 nezávislých datových souborů, na nichž odhadujeme neznámý parametr intervalem spolehlivosti, tak zhruba 95 intervalů bude hledaný parametr obsahovat a zhruba pět nikoli
- Příklad: výška **40** náhodně vybraných politiků
 - průměr: **175** cm



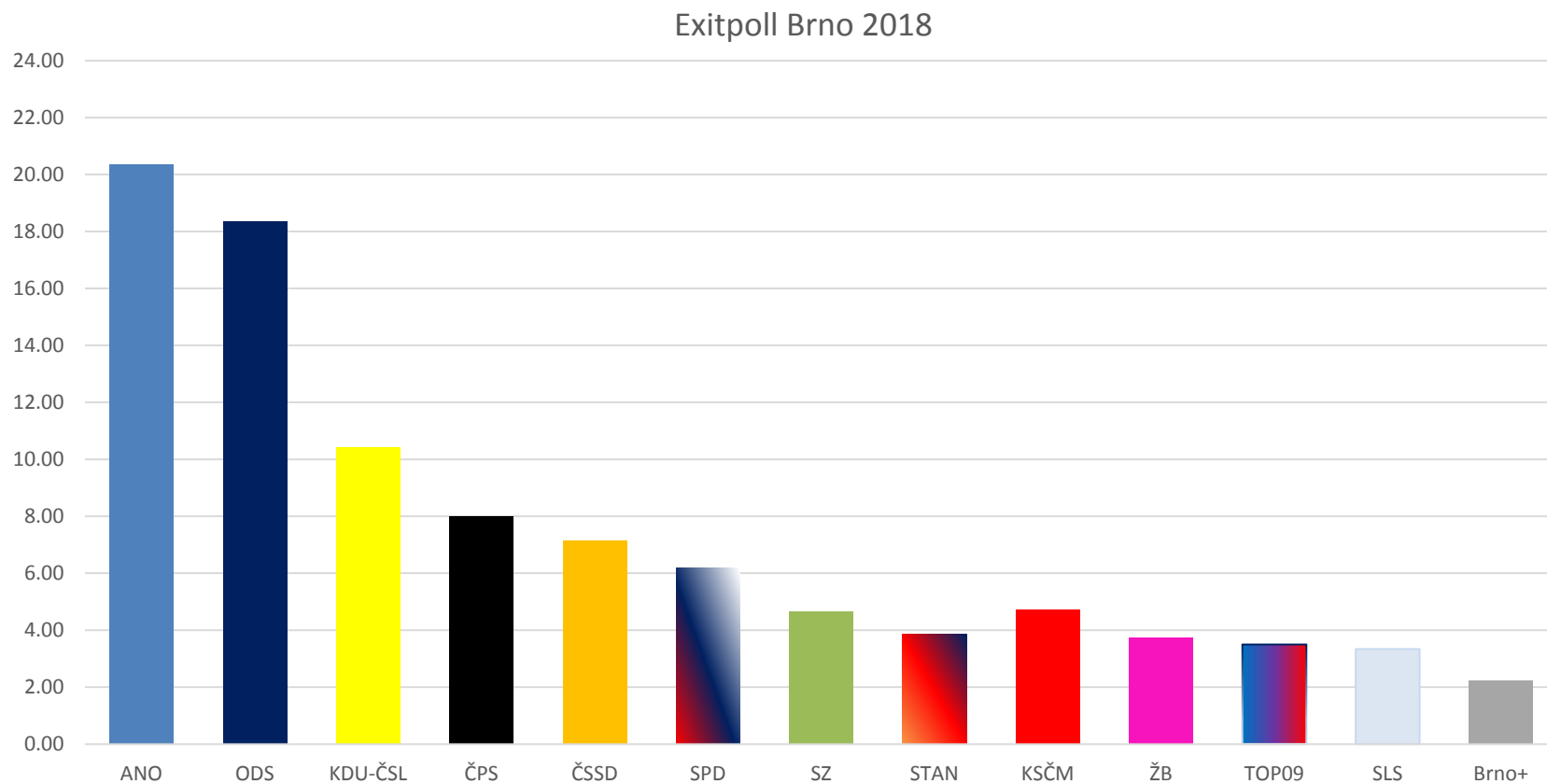
10 cm



$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

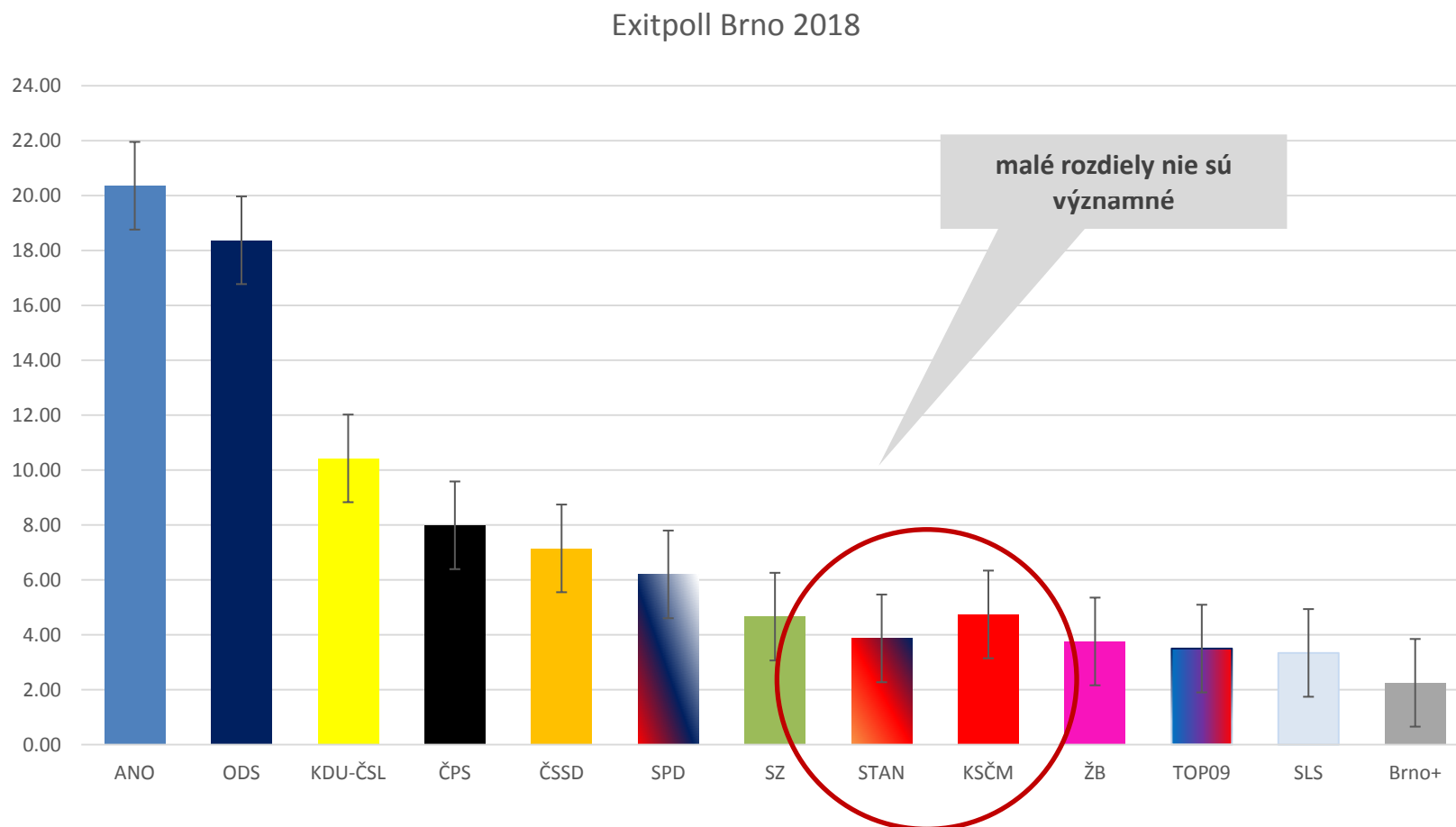
INTERVAL SPOLEHLIVOSTI

- exit poll – Magistrát Brno 2018



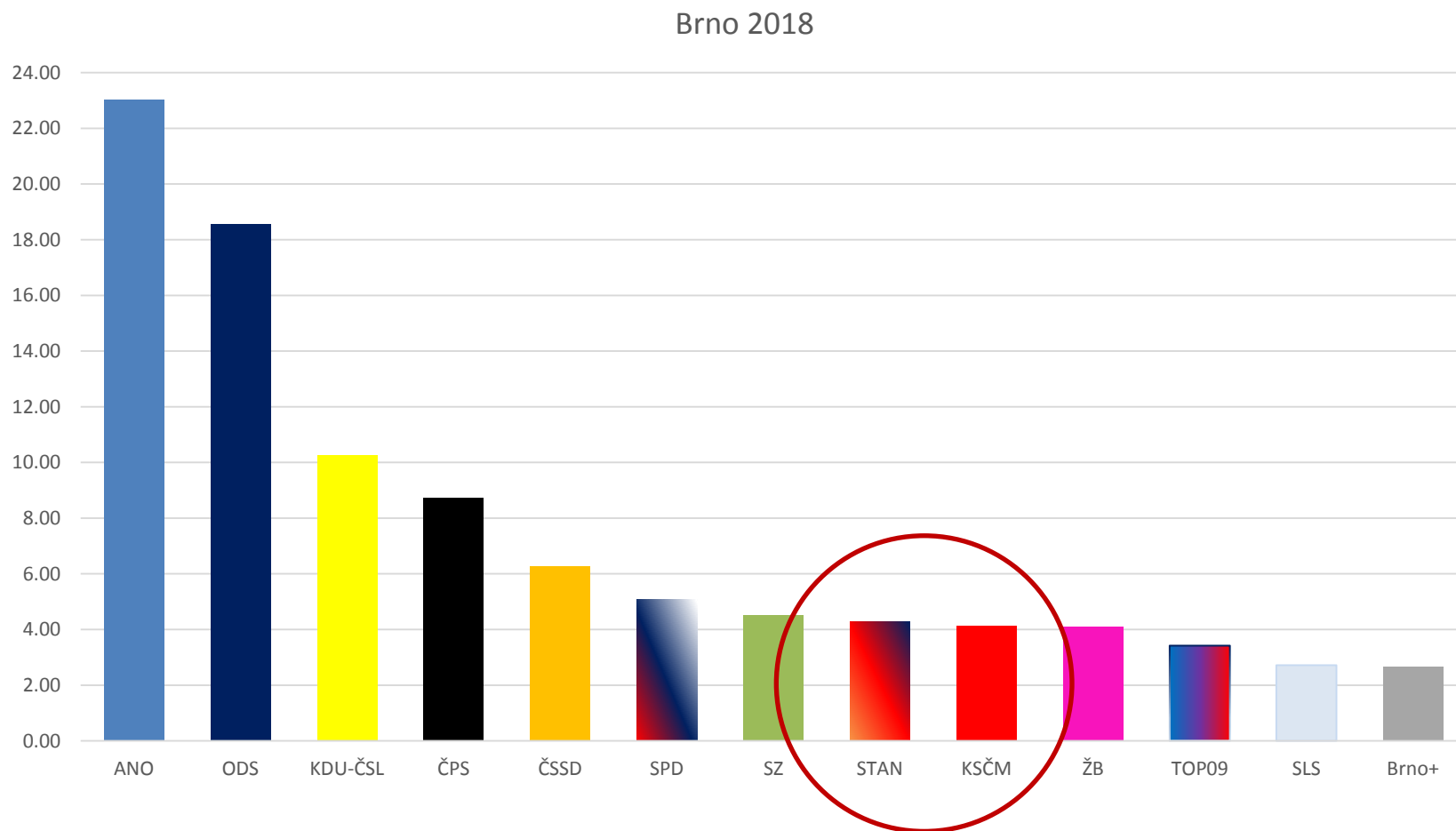
INTERVAL SPOLEHLIVOSTI

- exit poll – Magistrát Brno 2018



INTERVAL SPOLEHLIVOSTI

- reálné výsledky – Magistrát Brno 2018



VZTAHY MEZI PROMĚNNÝMI

VZTAHY MEZI PROMĚNNÝMI

- dvojrozměrná analýza – vztahy mezi dvěma proměnnými
- **kontingenční tabulka** (*crosstabs*) – základná vizualizace vztahu dvou statistických znaku (hodí se hlavně pro kategorické proměnné s malým množstvím kategorií)
- **nerozlišujeme mezi závislou a nezávislou p.**

řádky = hodnoty prvního znaku/proměnné (pohlaví)

sloupce = hodnoty druhého znaku/proměnné (dominantní ruka)

	Pravá	Ľavá	Spolu
Muž	43	9	52
Žena	44	4	48
Spolu	87	13	100

VZTAHY MEZI PROMĚNNÝMI

- typický brněnský volič ANO ?

		věk							Total
		18 - 21 let	22 - 30 let	31 - 40 let	41 - 50 let	51 - 50 let	61 - 70 let	71 let a více	
Volba strany	ODS	11	25	42	59	40	27	20	224
	CSSD	3	5	10	14	10	15	21	78
	STAN	3	16	26	13	6	3	4	71
	KSCM	1	2	4	0	11	7	15	40
	Pirati	22	77	51	18	9	11	4	192
	TOP09	12	38	36	33	18	10	6	153
	ANO	6	12	15	32	36	71	78	250
	KDU-CSL	6	14	17	23	16	12	28	116
	SPD	7	13	14	11	17	19	8	89
	ina strana	3	10	5	6	7	7	4	42
Total		78	238	249	218	178	186	191	1338

VZTAHY MEZI PROMĚNNÝMI

- typický brněnský volič ANO ?

Volba strany * vzdělání Crosstabulation

			vzdělání				Total
			základní	vyučený	maturita	vš	
Volba strany	ina strana	Count	2	5	16	19	42
		% within Volba strany	4,8%	11,9%	38,1%	45,2%	100,0%
		% within vzdělání	4,2%	3,8%	3,0%	3,1%	3,1%
SPD		Count	11	15	36	27	89
		% within Volba strany	12,4%	16,9%	40,4%	30,3%	100,0%
		% within vzdělání	22,9%	11,5%	6,7%	4,3%	6,7%
KDU-CSL		Count	0	7	38	71	116
		% within Volba strany	0,0%	6,0%	32,8%	61,2%	100,0%
		% within vzdělání	0,0%	5,3%	7,1%	11,4%	8,7%
ANO		Count	13	43	114	78	248
		% within Volba strany	27,1%	32,8%	21,2%	12,6%	100,0%
		% within vzdělání	27,1%	32,8%	21,2%	12,6%	18,5%

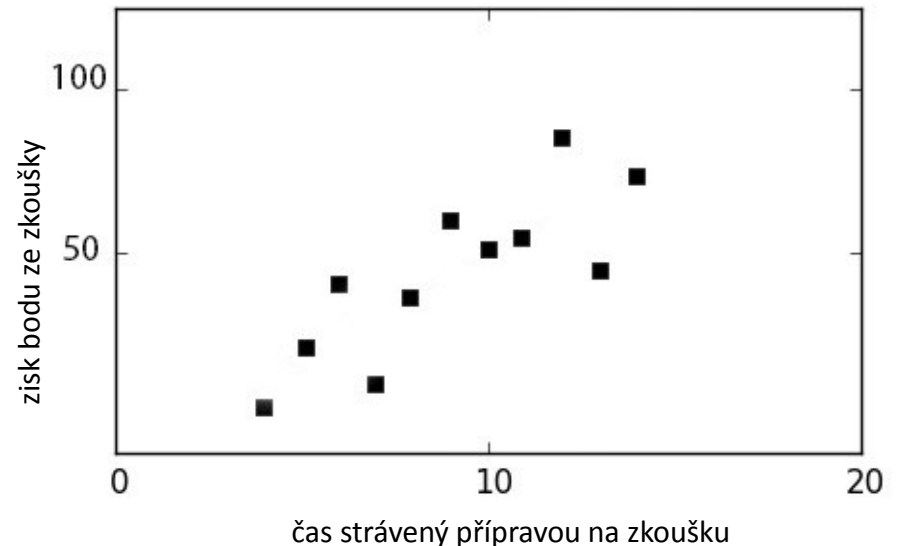
VZTAHY MEZI PROMĚNNÝMI

- typický brněnský volič ANO ?

			povolání							Total	
			zaměstnavatel	student	manuálně pracující	nemanuálně pracující	nezaměstnaný	rodičovská	důchodce		jiné
Volba strany	ina strana	Count	8	4	4	11	1	3	11	0	42
		% within Volba strany	19,0%	9,5%	9,5%	26,2%	2,4%	7,1%	26,2%	0,0%	100,0%
		% within povolání	4,0%	2,8%	3,4%	2,3%	9,1%	6,4%	3,3%	0,0%	3,1%
SPO		Count	1	0	1	0	0	0	0	0	2
		% within Volba strany	50,0%	0,0%	50,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
		% within povolání	0,5%	0,0%	0,8%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%
SPD		Count	13	5	20	20	1	4	26	0	89
		% within Volba strany	14,6%	5,6%	22,5%	22,5%	1,1%	4,5%	29,2%	0,0%	100,0%
		% within povolání	6,5%	3,5%	16,8%	4,2%	9,1%	8,5%	7,7%	0,0%	6,6%
KDU-CSL		Count	10	9	7	42	1	8	38	1	116
		% within Volba strany	8,6%	7,8%	6,0%	36,2%	0,9%	6,9%	32,8%	0,9%	100,0%
		% within povolání	5,0%	6,3%	5,9%	8,8%	9,1%	17,0%	11,3%	20,0%	8,7%
ANO		Count	22	12	24	51	2	6	133	0	250
		% within povolání	11,1%	8,5%	20,2%	10,6%	18,2%	12,8%	39,6%	0,0%	18,7%

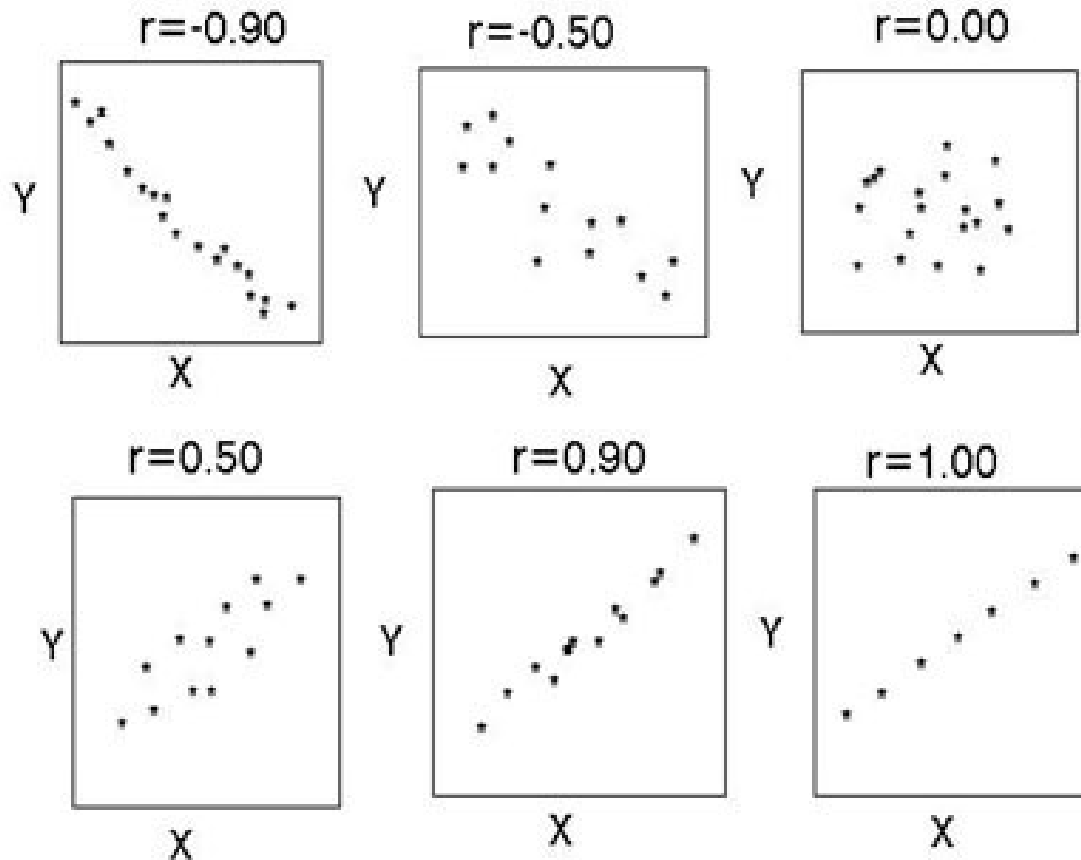
VZTAHY MEZI PROMĚNNÝMI

- kontingenční tabulka nám ale neřekne nic o (síle) vztahu
- **existuje korelace? proměnné spolu souvisí?**
 - statistický ukazovatel vztahu mezi 2 proměnnými
 - změny v jedné proměnné jsou doprovázeny změnami v druhé proměnné
 - vzájemná asociace mezi proměnnými je **lineární**
- **korelační koeficient (r)**
 - síla a směr vztahu
 - hodnoty $<-1,1>$, 0 = absence vztahu



VZTAHY MEZI PROMĚNNÝMI

- korelační koeficient

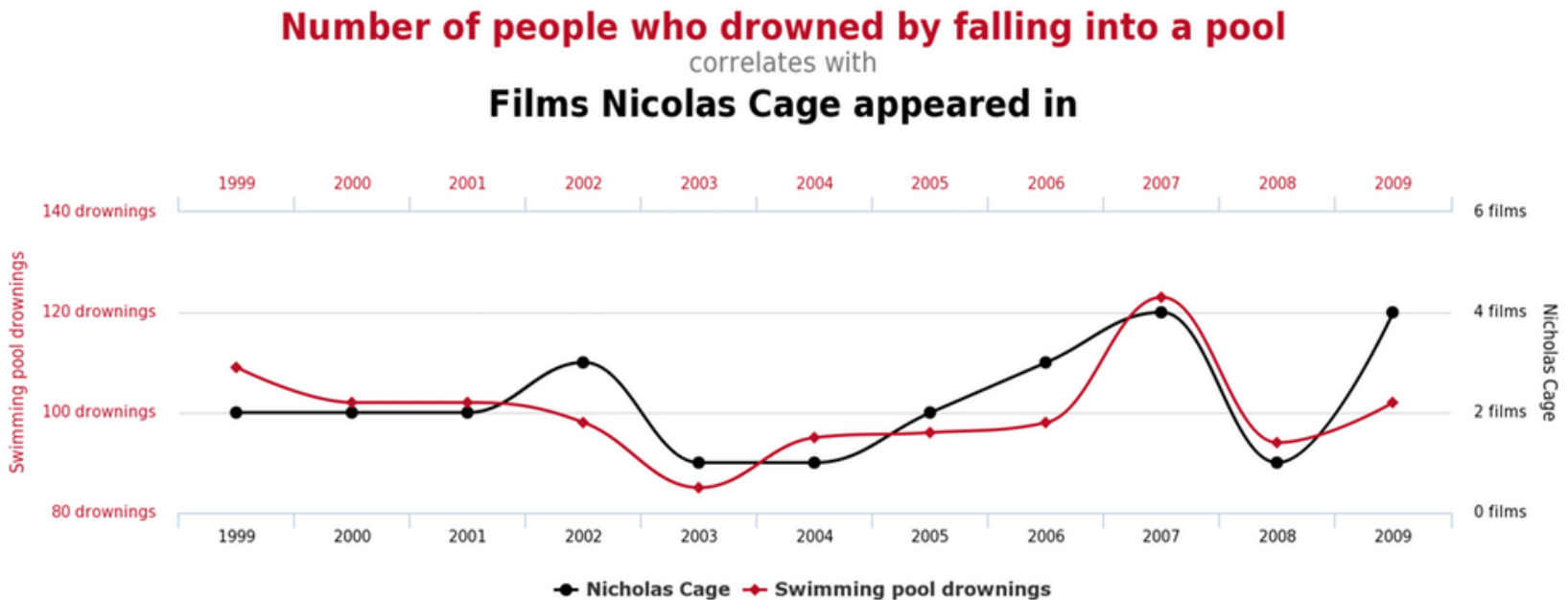


**GUESS THE
CORRELATION**

VZTAHY MEZI PROMĚNNÝMI

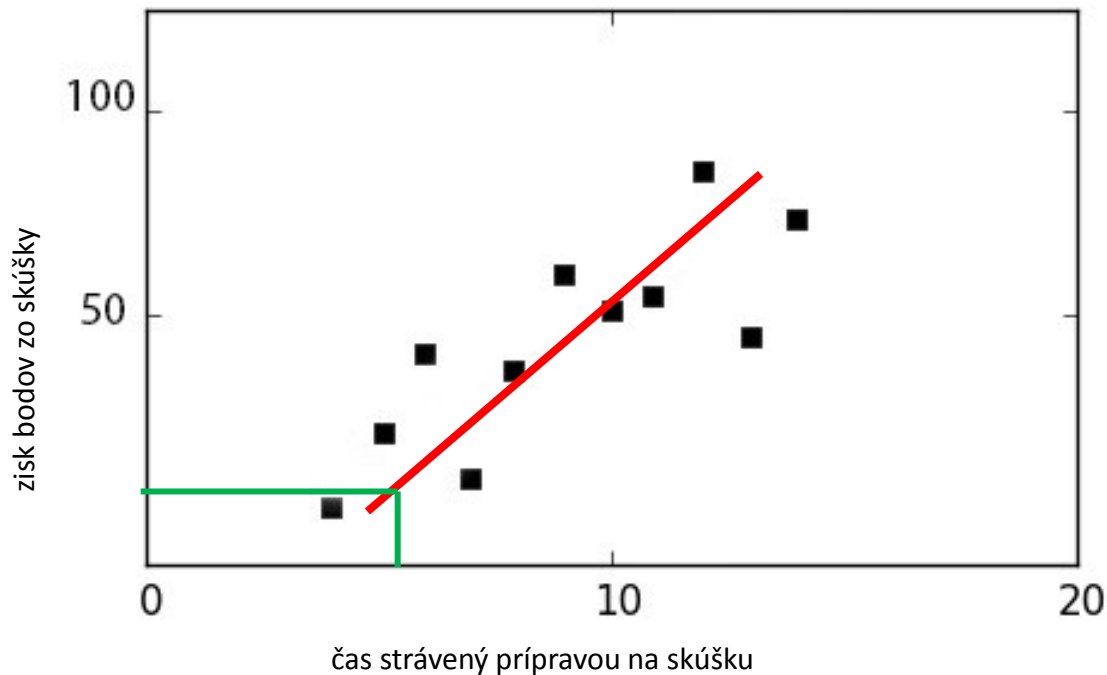
- **korelace není kauzalita!** ... korelace **neznamená**, že jedna věc ovlivňuje druhou (můžou existovat jiné důvody vzájemné korelace).

závislá a ~~nezávislá~~ proměnná



VZTAHY MEZI PROMĚNNÝMI

- **závislost** 2 a více proměnných → **regrese**
- **vplyv** nezávislé X na závislou Y
- **predikce**



VZTAHY MEZI PROMĚNNÝMI

- lineární **regrese**
 - závislá proměnná musí být kardinální (intervalová, poměrová)

příklad:

Co může ovlivňovat čas strávený denně na internetu?

Regresní model

data: European Social Survey 2016

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,303 ^a	,092	,089	157,562

vysvětlující síla modelu

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	253,722	16,233		15,630	,000
	Age of respondent, calculated	-3,040	,280	-,285	-10,847	,000
	Highest level of education	,148	,026	,142	5,596	,000
	Trust in politicians	-4,825	4,178	-,067	-1,155	,248
	Trust in political parties	-2,676	4,273	-,036	-,626	,531
	How interested in politics	17,780	5,834	,082	3,048	,002

regresní koeficient

statistická významnost
 $p < 0,05$

a. Dependent Variable: Internet use, how much time on typical day, in minutes

Report výsledků



VIZUALIZACE DAT

VIZUALIZACE DAT

- prezentace výstupů
- posouzení charakteru dat
- ulehčuje porozumět analýzám
- „jeden obrázek řekne víc než tisíc slov“
- graf -> příběh
- **velmi silný nástroj**



VIZUALIZACE DAT - úlohy

- Definujte svoje publikum
- Vytvořte správu, kterou chcete komunikovat
- Definujte povahu správy

A good chart:

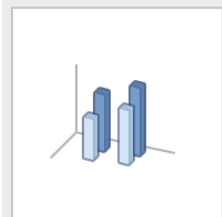
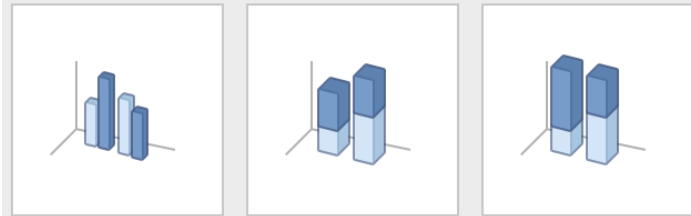
- grabs the reader's attention;
 - presents the information simply, clearly and accurately;
 - does not mislead;
 - displays the data in a concentrated way (e.g. one line chart instead of many pie charts);
 - facilitates data comparison and highlights trends and differences;
 - illustrates messages, themes or storylines in the accompanying text.
- United nations economic commission for europe – „Making data meaningful“

Výběr typu grafu

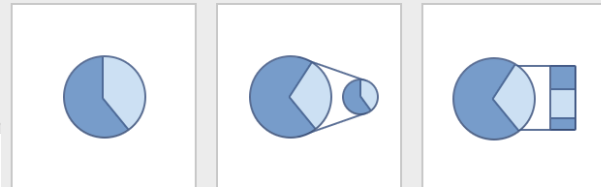
2D stĺpcový



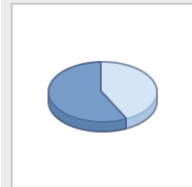
3D stĺpcový



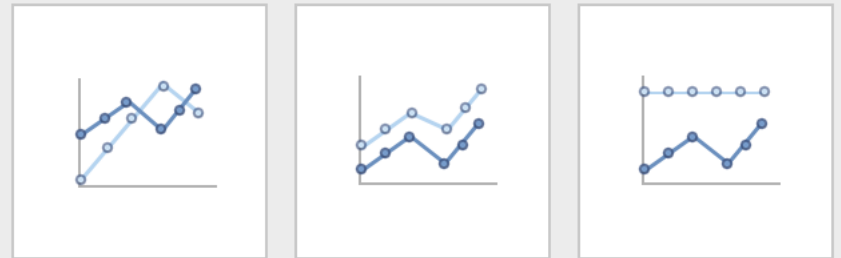
2D koláčový



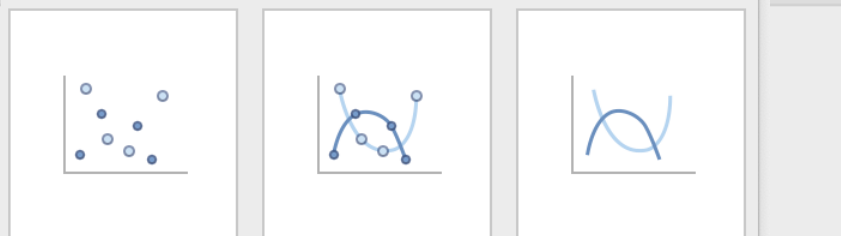
3D koláčový



2D čiarový

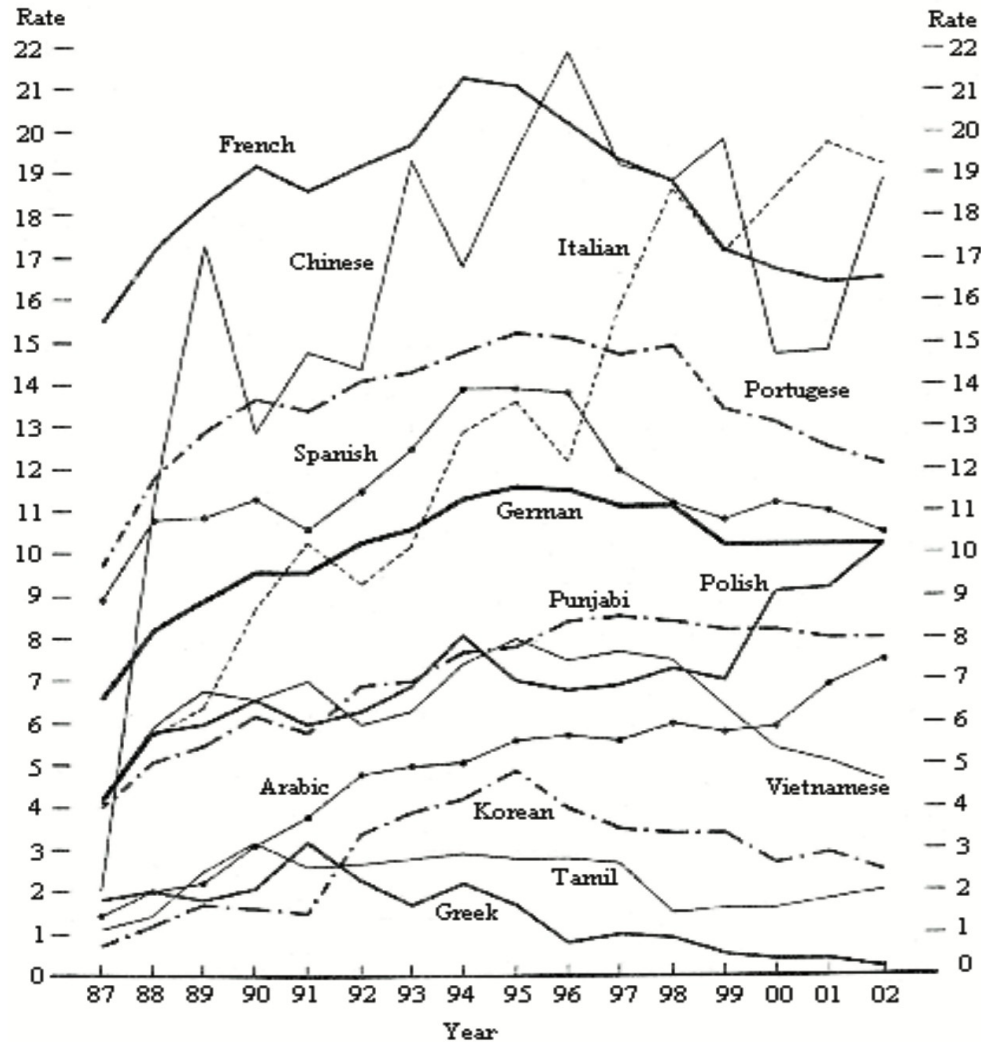


Bodový

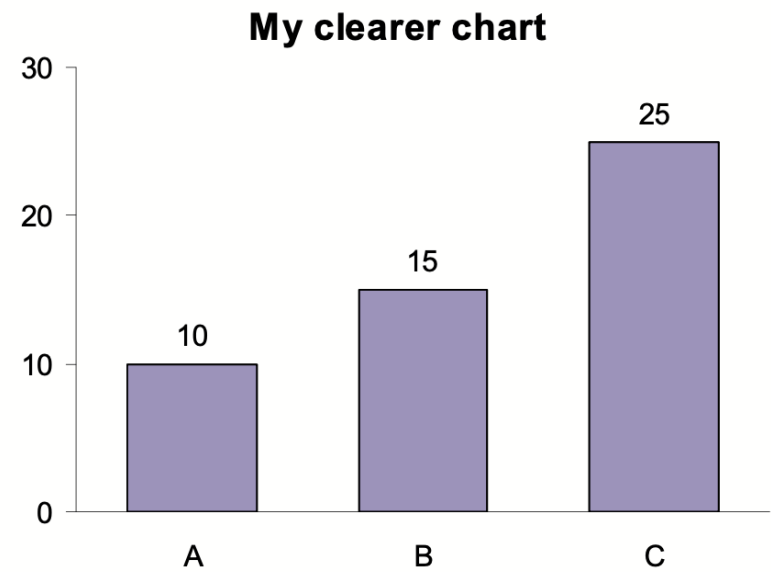
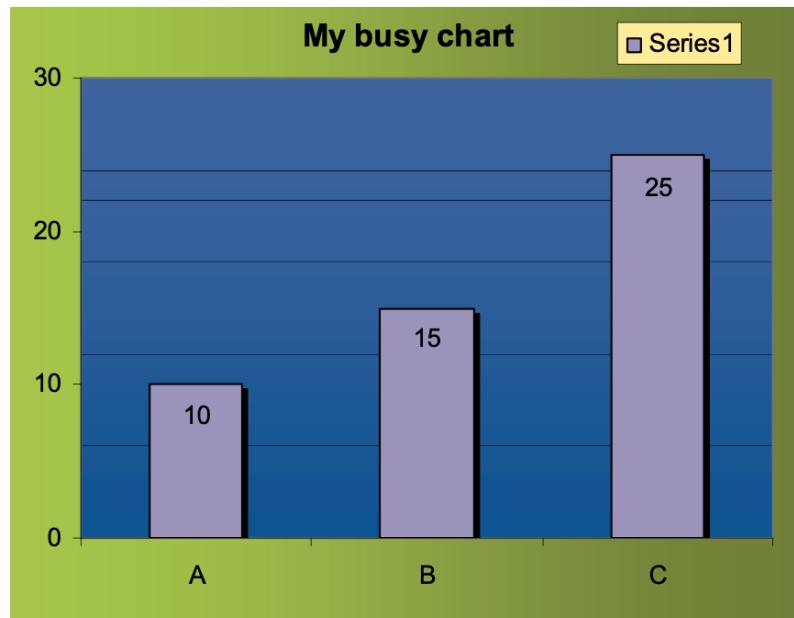


Jasná message?

Number of students taking English as a second language at West High School, by first language spoken, 1987 to 2002

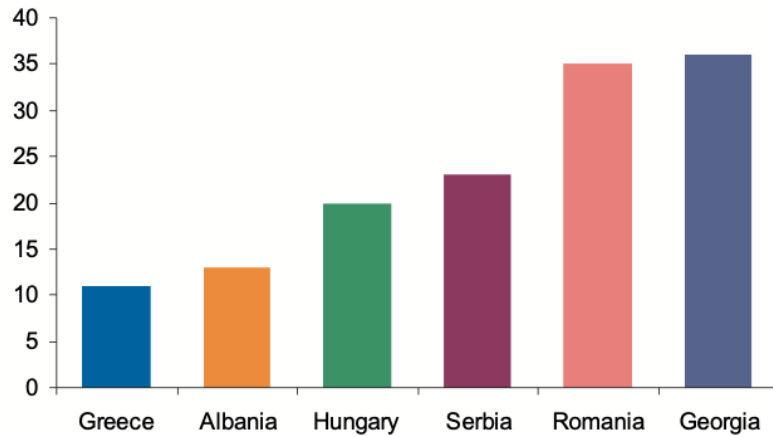


Zbytečnosti?

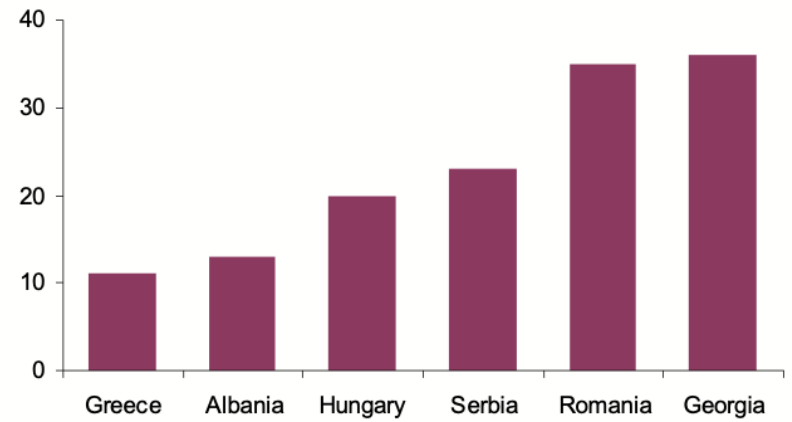


Zbytečnosti?

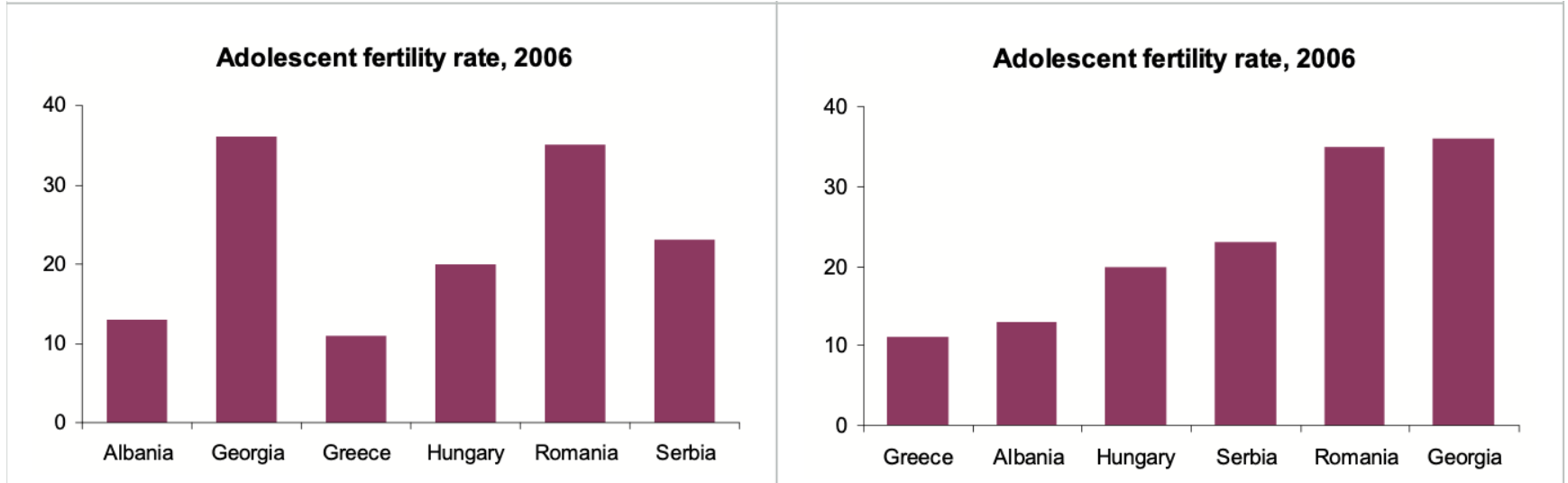
Adolescent fertility rate, 2006



Adolescent fertility rate, 2006

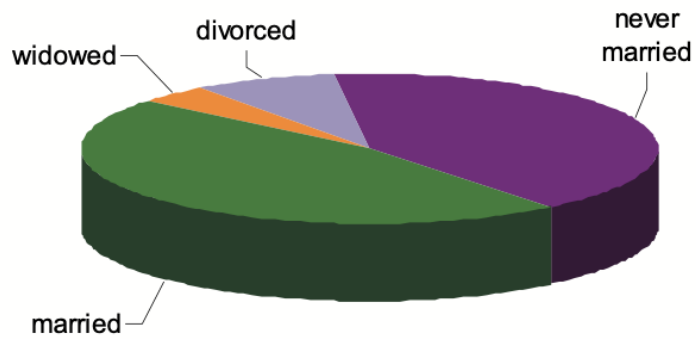


Zjednodušující?

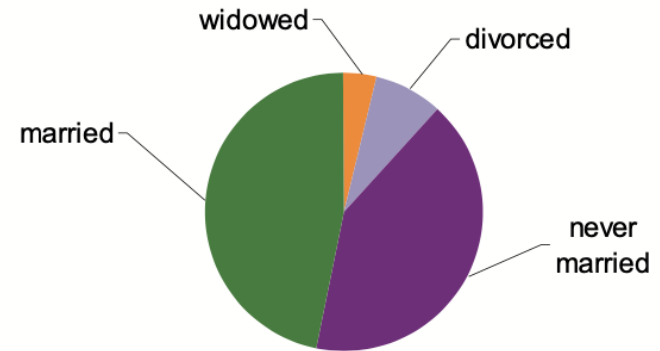


Zjednodušující?

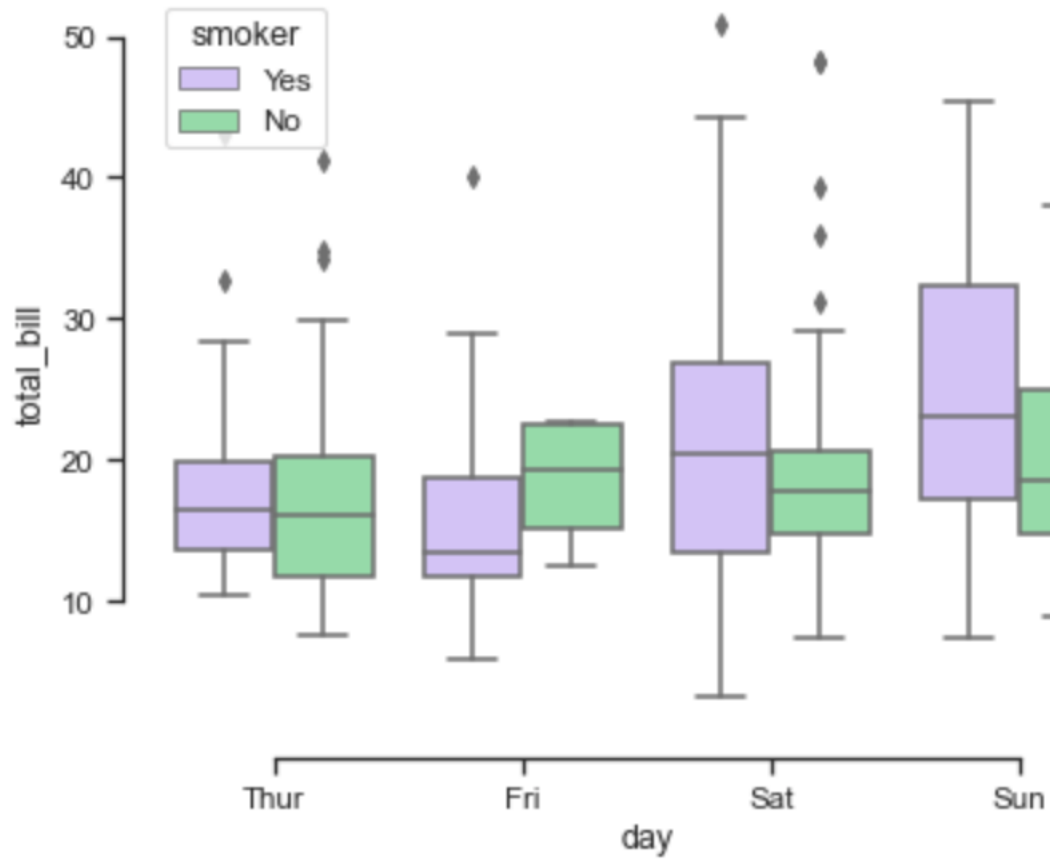
Population aged 18+ by legal marital status in Iceland, 2004



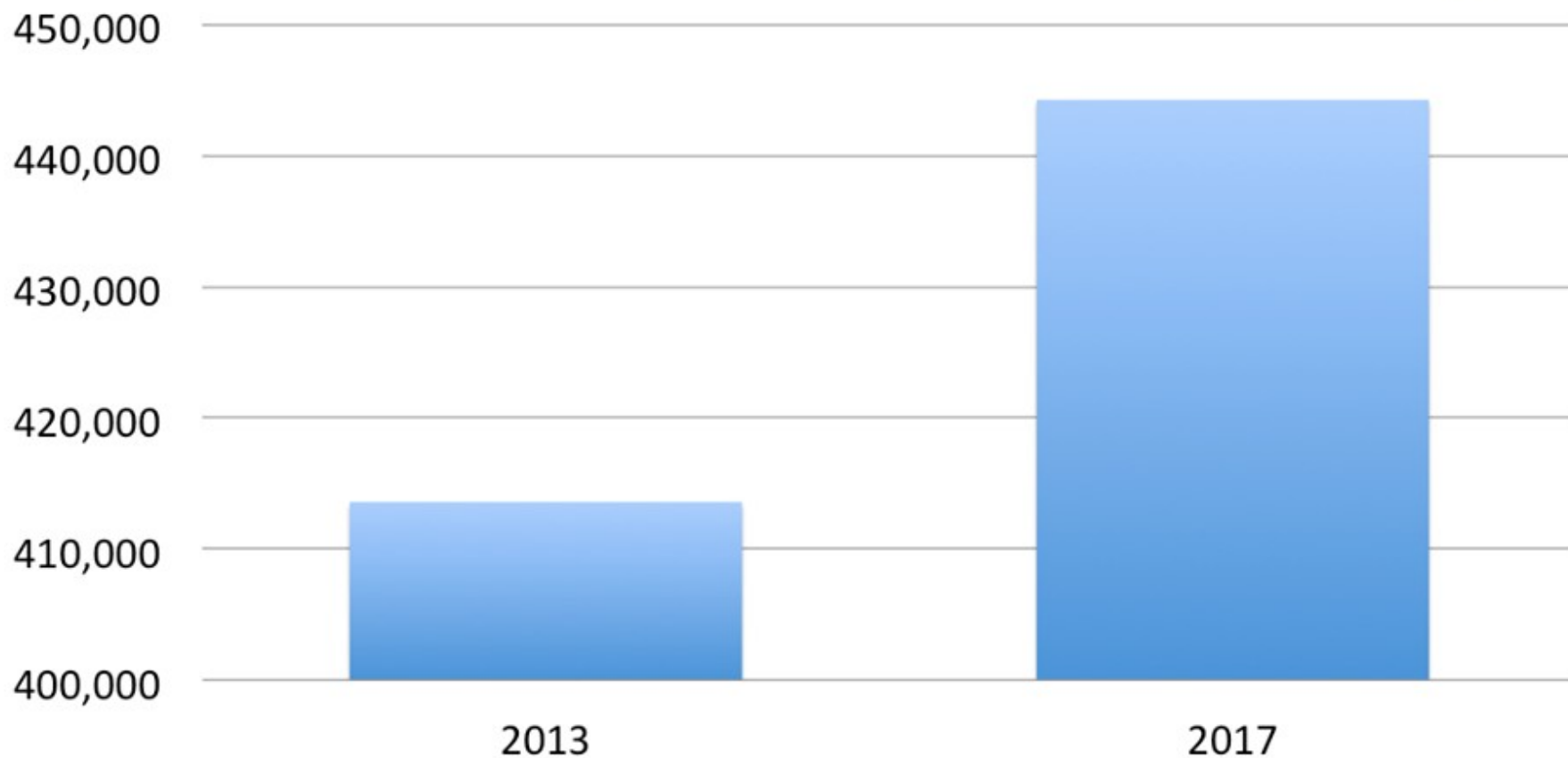
Population aged 18+ by legal marital status in Iceland, 2004



Kognitivně náročný?



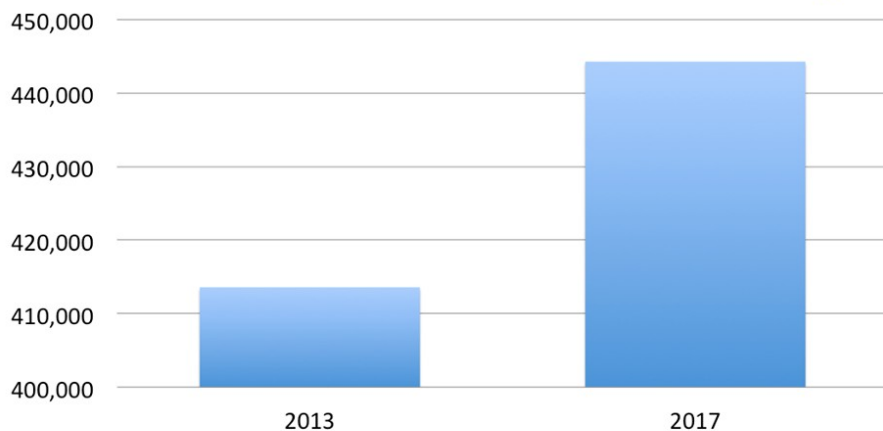
Počet státních zaměstnanců



Zdroj: Ministerstvo financí; Návrh státního rozpočtu 2017



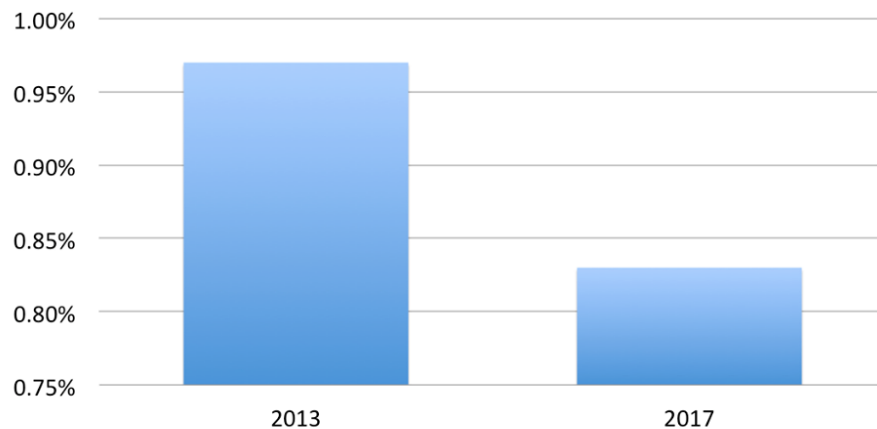
Počet státních zaměstnanců



Zdroj: Ministerstvo financí; Návrh státního rozpočtu 2017



Výdaje na vědu, výzkum a inovace k HDP v %



Zdroj: Ministerstvo financí; Návrh státního rozpočtu 2017



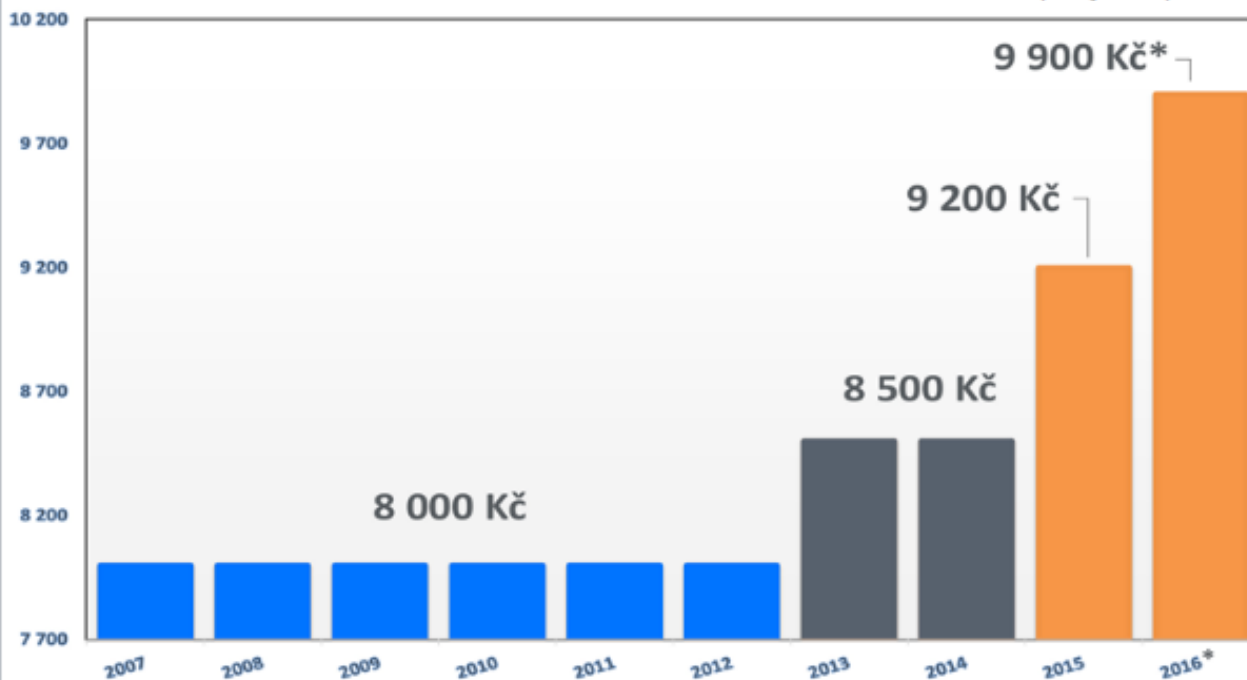
ODS - Občanská demokratická strana ✓

Pokud se člověk podívá na levou část grafu, tak přece stejně jako Vy vidí, kde graf začíná. Cílem grafu je ukázat trend.

Like · Reply · 30 September 2016 at 17:24

Vývoj minimální mzdy v ČR v Kč od roku 2007

(zdroj MPSV)



Úřad vlády ČR @strakovka · Aug 20

@SlavekSobotka



5



8



Úřad vlády ČR @strakovka · Jul 29

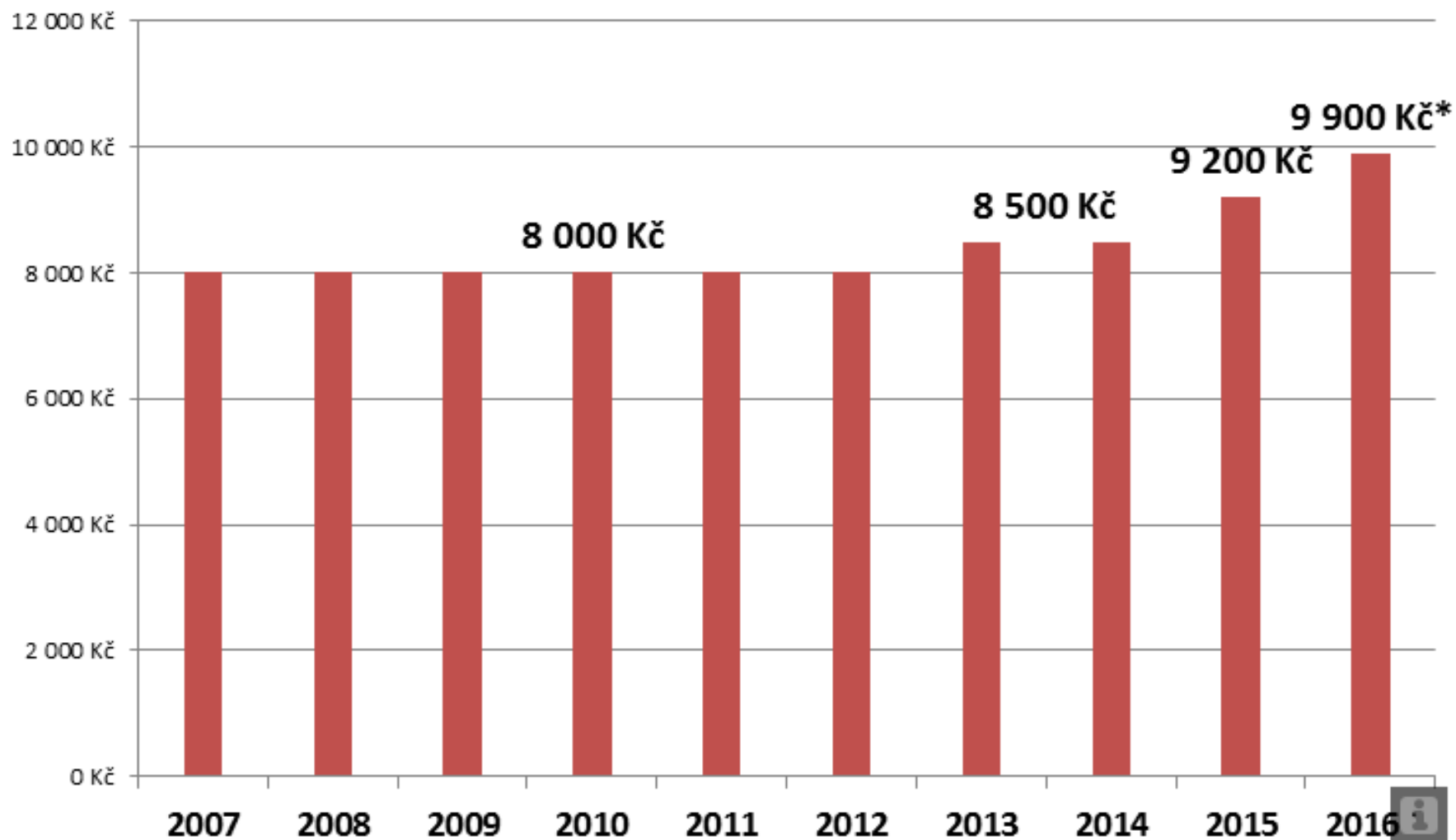
Včerejší jednání vlády skončilo po 16 hodinách, výsledky zde:
[vlada.cz/czimedia_contr](#)



Robert Casensky
@RCasensky

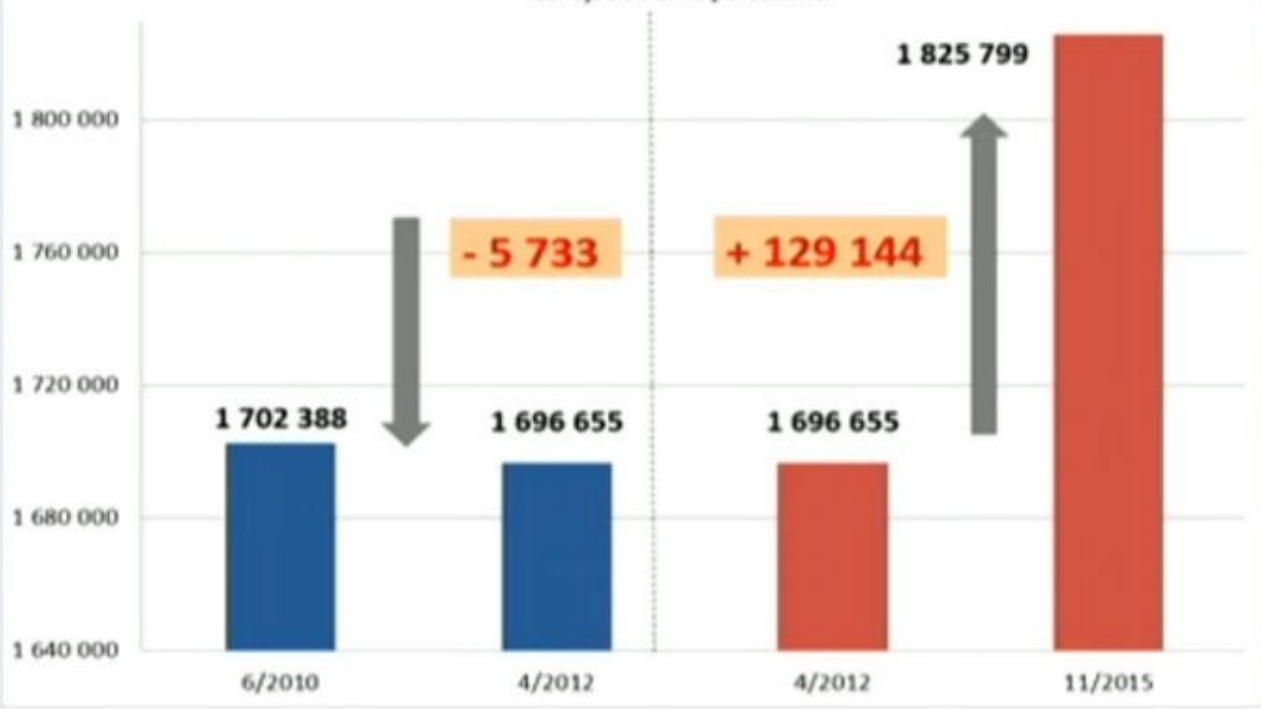


Vývoj minimální mzdy v ČR v Kč od roku 2007 (zdroj MPVS)



Počet zamestnancov počas vlády I. Radičovej a R. Fica

zdroj: Sociálna poisťovňa



TA3

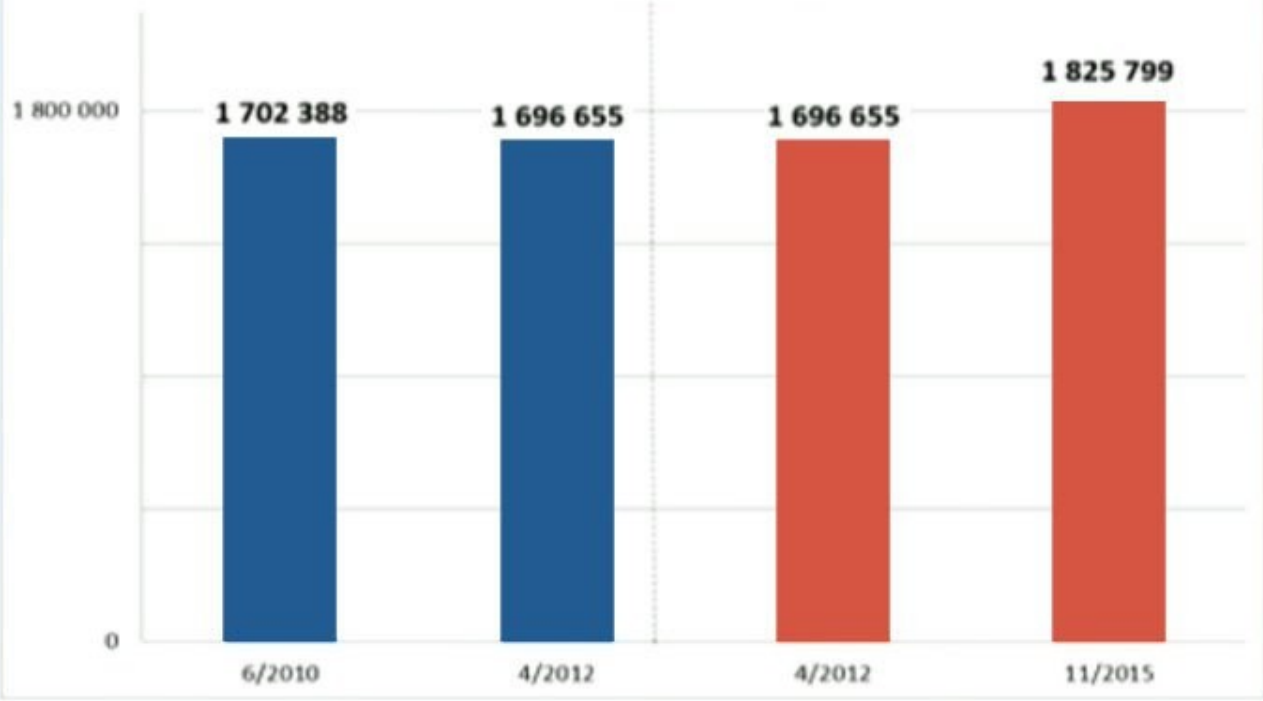
ITTER.COM/TVTA3. DOMÁCE = SESTRY BUDÚ ROKOVAŤ SO Š

11:42

NAŽIVO

Počet zamestnancov počas vlády I. Radičovej a R. Fica

zdroj: Sociálna poisťovňa



TA3

ITTER.COM/TVTA3. DOMÁCE = SESTRY BUDÚ ROKOVAŤ SO Š

Ďakujeme!

3,4 %

júl 2023

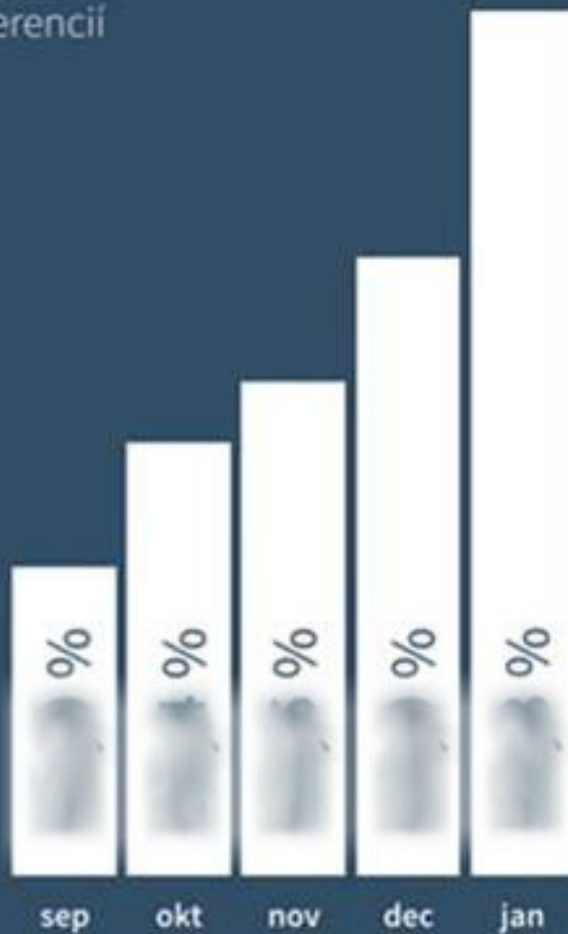
4,1 %

august 2023

**Sme odhodlaní bojovať
za vás aj naďalej!**

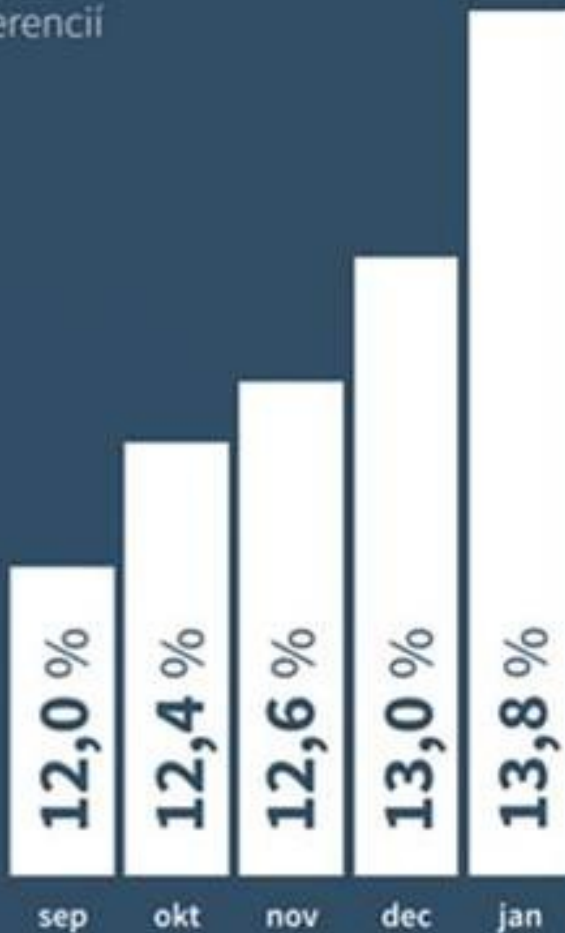
agentúra Polis - prieskum preferencií

#siet'



agentúra Polis - prieskum preferencií

#siet'

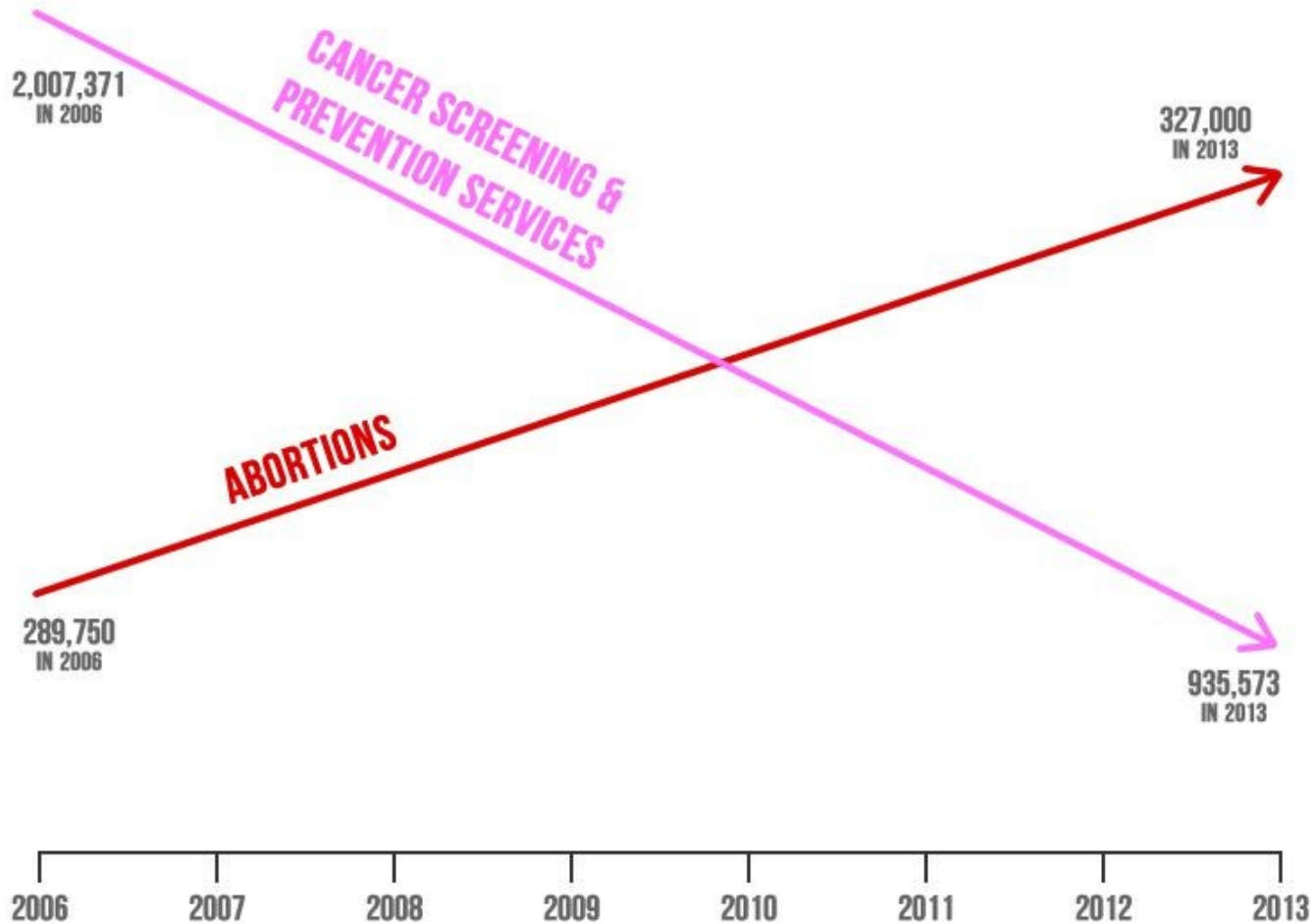


agentúra Polis - prieskum preferencií

#siet'



PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN



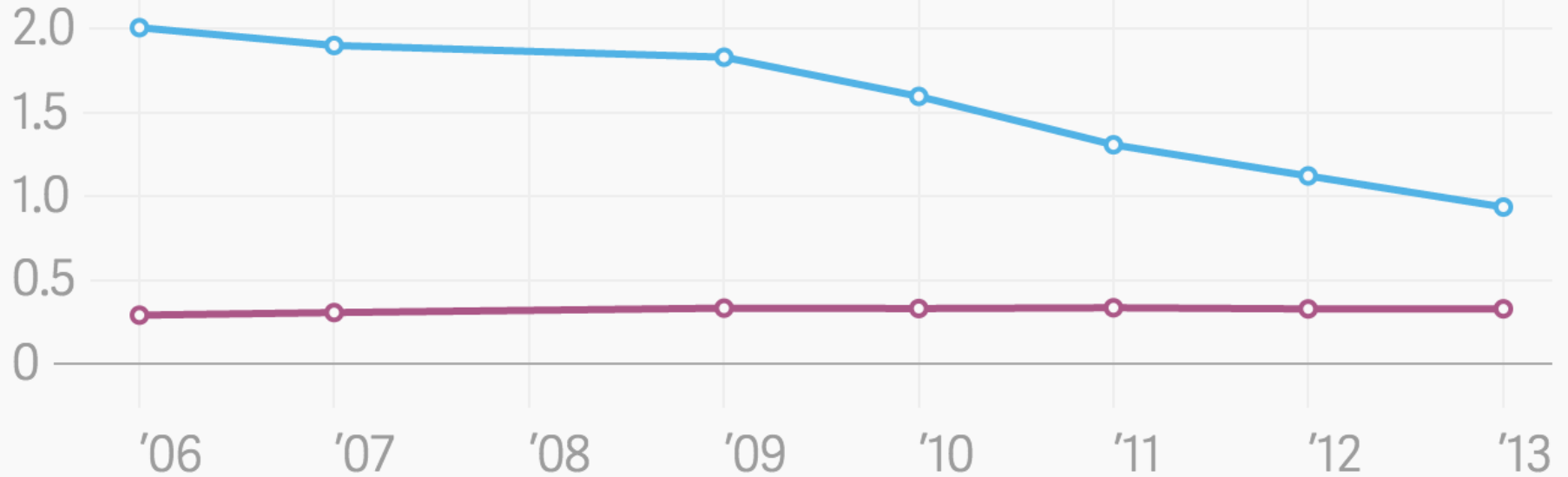
SOURCE: AMERICANS UNITED FOR LIFE

Planned Parenthood services

■ Abortion procedures

■ Cancer screening / preventative services

2.5 million



UNEMPLOYMENT RATE

UNDER PRESIDENT OBAMA



2011

SOURCE: BUREAU OF LABOR STATISTICS



WAR AND A TROOP WITHDRAWAL AT THE E NAS FUT 2,292.50

UNEMPLOYMENT RATE

UNDER PRESIDENT OBAMA



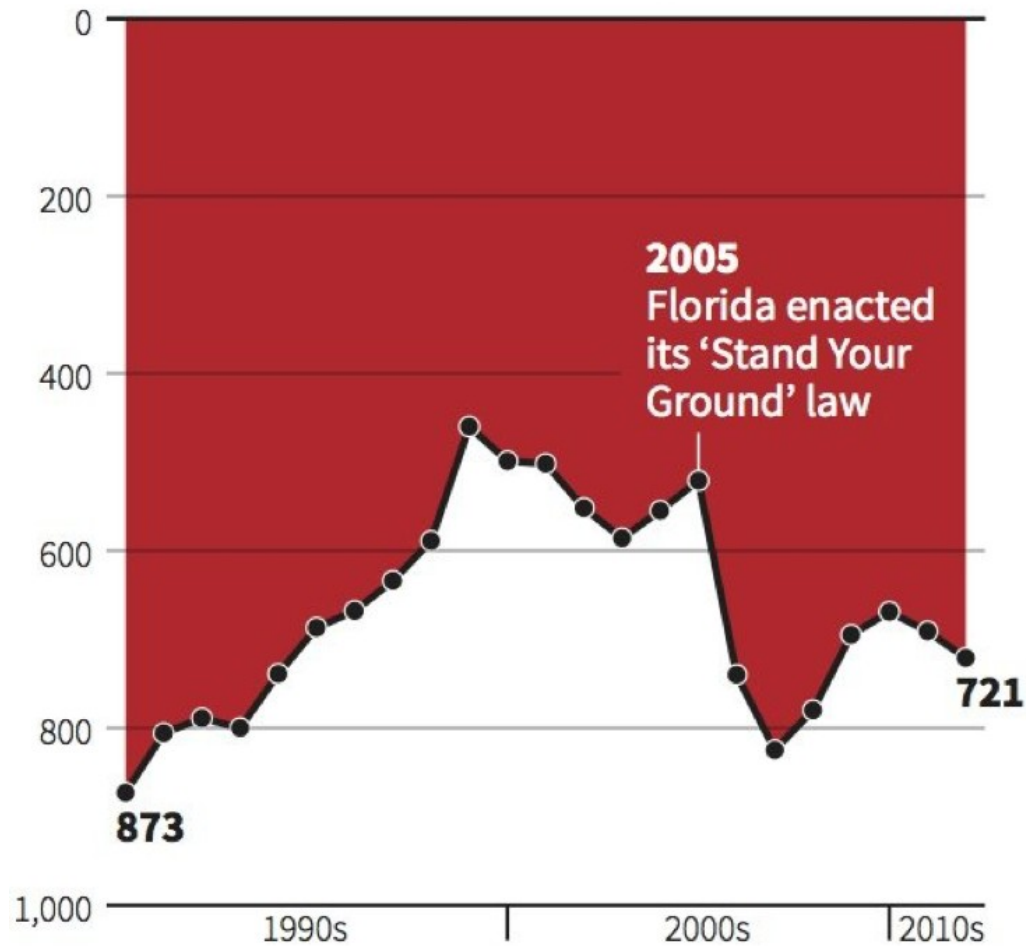
2011

SOURCE: BUREAU OF LABOR STATISTICS

WAR AND A TROOP WITHDRAWAL AT THE E NAS FUT 2,292.50

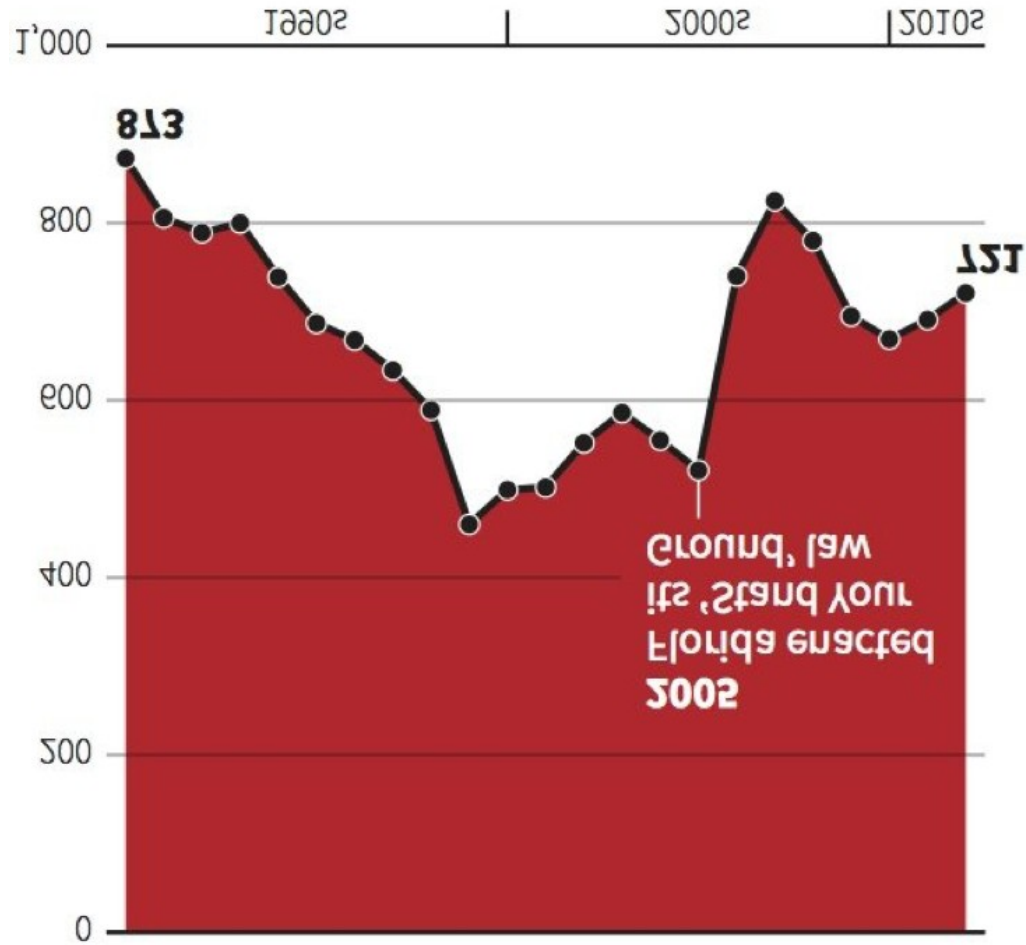
Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

SOURCE: Florida Department of Law Enforcement



Number of murders committed using firearms

gun deaths in Florida

01

PENÍZE ZA VOLBY



34 mil. Kč



34 mil. Kč



16 mil. Kč



13 mil. Kč



12 mil. Kč



8 mil. Kč

Zdroj: ČTK

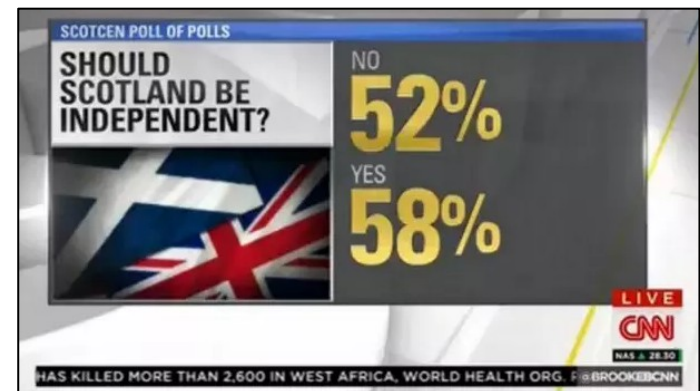
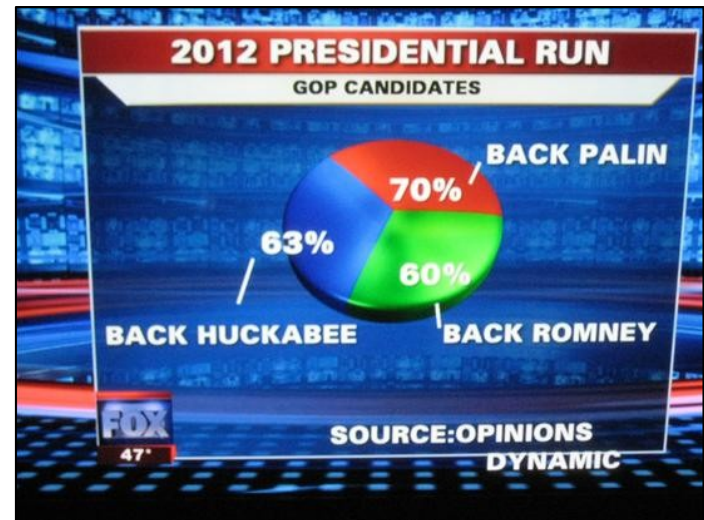


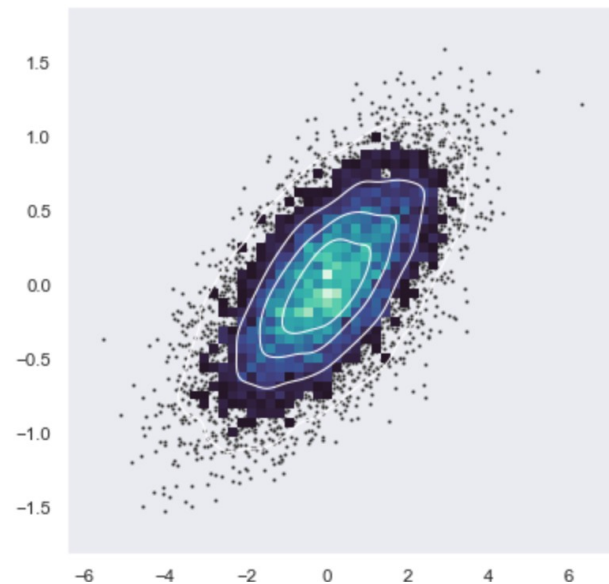
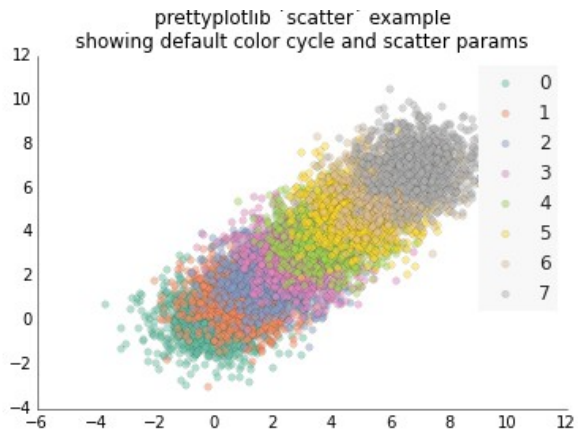
VIZUALIZACE DAT

- i jemná změna může mít zásadní dopad

Nejčastěji:

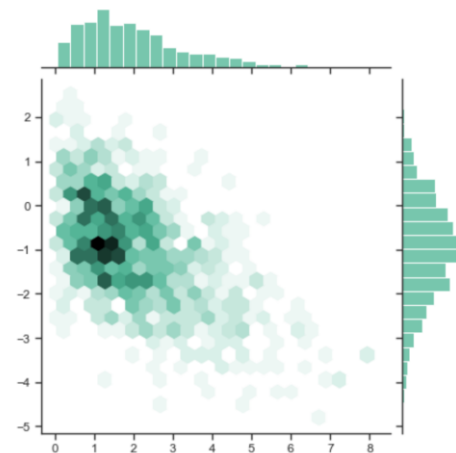
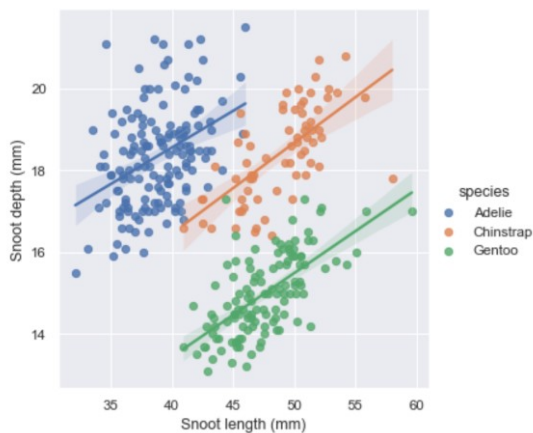
- chybějící východiskový bod – posunutá Y-os
- dvě Y-osi v jednom grafu
- vynechané osy a jejich popisy
- chybně nakreslený graf





Vizualizace „Život je jako bonboniéra“

Forrest Gump





Shrnutí

- **Sběr** dat
- **Analýza** dat
- **Interpretace** dat

- Střední hodnoty
- Normální distribuce
- Populace vs. Vzorek
- Vizualizace