Popisná statistika
Kvalitativní proměnná

Martina Litschmannová

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

Typy statistických znaků (proměnných)

Nominální
- varianty jsou ve formátu text nebo číselný kód
- o každých dvou variantách lze říci, zda jsou různé
- např. škola, fakulta, obor, výrobce, …
- Další dělení: dichotomické (alternativní), vícekategoriální (množné)

Ordinální (pořadová)
- varianty jsou ve fomátu text, datum nebo číslo
- u každých dvou variant lze stanovit jejich pořadí
- např. úroveň vzdělání, známka (A, B, …, E), úroveň spokojenosti, …

Kvalitativní

Intervalové (rozdílové)
- varianty jsou v číselném formátu
- u každých dvou variant lze určit jejich pořadí a rozdíl
- např. teplota ve °C, chyba měření, …

Poměrové
- varianty jsou v číselném formátu (pouze kladná čísla + nulový bod)
- u každých dvou variant lze určit jejich pořadí, rozdíl a podíl (poměr)
- např. teplota v K, velikost chyby měření, …

Kvantitativní
(numerické, kardinální)
Další dělení: diskrétní, spojité

Popisná statistika
aneb
Jak efektivně popsat a vizualizovat data

Kvalitativní znak

Popisná statistika – Kvalitativní znak

Časová značka	Pohlaví	Výška (cm)	Váha (kg)	Přivyděláváte si v rámci prezenčního studia na brigádách?	Jak často brigádu máte?	Jak byste svou brigádu charakterizoval(a)?	Kolik času týdně obvykle věnujete brigádě?	Kolik času týdně obvykle věnujete studiu?
ID	pohlaví	výška (cm)	váha (kg)	brigáda	frekvence brigády	charakteristika brigády	čas věnovaný brigádě (h/týden)	čas věnovaný studiu (h/týden)
1.4.2016 10:38	muž	180	70	ano	každý pracovní den	praxe v oboru během studia	20	15
1.4.2016 10:41	muž	186	85	ano	nepravidelně	kancelářská práce a na ní navazující práce manuální při realizaci projektů	30	20
1.4.2016 10:41	muž	172	75	ano	nepravidelně	praxe v oboru během studia	5	36
1.4.2016 10:45	žena	166	56	ano	Různě, 2-3 týdně	Hlídání dětí	12	10
1.4.2016 10:52	žena	188	70	ano	3 dny v tydnu	praxe v oboru během studia	24	26

Respondent (proband) – označení statistické jednotky v dotazníkovém šetření

Popište strukturu datového souboru v závislosti na pohlaví respondentů.

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Varianta znaku x_i	**Absolutní četnost** n_i	**Relativní četnost** p_i
x_1	n_1	$p_1=n_1/n$
x_2	n_2	$p_2=n_2/n$
x_k	n_k	$p_k=n_k/n$
Celkem:	$n_1+n_2+\ldots+n_k=n$	1

- **Modus** (varianta, které znak nabývá s nejvyšší četností)

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost**
muž	66	0,776470588
žena	19	0,223529412
Celkem:	85	1,000000000

Jak zaokrouhlovat relativní četnosti?

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost**
muž	66	0,776470588
žena	19	0,223529412
Celkem:	85	1,000000000

Jak zaokrouhlovat relativní četnosti?

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	77,6470588
žena	19	22,3529412
Celkem:	85	100,0000000

Pozor na zaokrouhlovací chybu!

Jak zaokrouhlovat relativní četnosti?

1,0 % … 0,85 osob
0,1 % … 0,085 osob

Popisná statistika – Kvalitativní znak

- **Tabulka četnosti**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	78
žena	19	22
Celkem:	85	100

Pozor na zaokrouhlovací chybu!

Součet musí být 100 %!

Jak zaokrouhlovat relativní četnosti?

1,0 % … 0,85 osob
0,1 % … 0,085 osob

Příklad demonstrující problém zaokrouhlovací chyby

- **Tabulka četností**

TABULKA ROZDĚLENÍ ČETNOSTI		
Typ pasažéra	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,37864
Žena	85	41,26214
Dítě	44	21,35922
Celkem:	206	100,00000

Jak zaokrouhlovat relativní četnosti?

1,0 % … 2,06 osob
0,1 % … 0,206 osob

Příklad demonstrující problém zaokrouhlovací chyby

- **Tabulka četností**

TABULKA ROZDĚLENÍ ČETNOSTI		
Typ pasažéra	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,4
Žena	85	41,3
Dítě	44	21,4
Celkem:	206	100,1

POZOR na zaokrouhlovací chybu!

Součet musí být 100 %!

Jak zaokrouhlovat relativní četnosti?

1,0 % … 2,06 osob
0,1 % … 0,206 osob

Příklad demonstrující problém zaokrouhlovací chyby

- **Tabulka četností**

TABULKA ROZDĚLENÍ ČETNOSTI		
Typ pasažéra	**Absolutní četnosti**	**Relativní četnosti (%)**
Muž	77	37,4
Žena	85	41,3
Dítě	44	21,3
Celkem:	206	100,0

POZOR na zaokrouhlovací chybu!

Dopočet do 100 %!

Jak zaokrouhlovat relativní četnosti?

1,0 % … 2,06 osob
0,1 % … 0,206 osob

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	?	78
žena	?	22
Celkem:	85	100

Relativní četnosti používejte vždy pouze jako doplněk absolutních četností!

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	78
žena	19	22
Celkem:	85	100

Určete modus pohlaví.

Popisná statistika – Kvalitativní znak

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	78
žena	19	22
Celkem:	85	100

Určete modus pohlaví.

modus = muž

Mezi respondenty převažovali muži.

Popisná statistika – Kvalitativní znak (nominální)

- **Tabulka četností**

Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	78
žena	19	22
Celkem:	85	100

Popisná statistika – Kvalitativní znak (ordinální)

- **Tabulka četností**

Tabulka četností				
Velikost	**Absolutní četnost**	**Relativní četnost (%)**	**Kumulativní četnost**	**Kumulativní rel. četnost (%)**
S	66	60,0	66	60,0
M	19	17,3	85	77,3
L	15	13,6	100	90,9
XL	10	9,1	110	100,0
Celkem:	110	100,0	---	---

- Tabulku četností rozšiřujeme o kum. četnosti a kum. relativní četnosti.

Kumulativní četnost (kum. rel. četnost) je postupně načítaná četnost (rel. četnost) jednotlivých vzestupně uspořádaných variant ordinálního znaku.

Popisná statistika – Kvalitativní znak (ordinální)

- **Tabulka četností**

Tabulka četností				
Velikost	**Absolutní četnost**	**Relativní četnost (%)**	**Kumulativní četnost**	**Kumulativní rel. četnost (%)**
S	66	60,0	66	60,0
M	19	17,3	85	77,3
L	15	13,6	100	90,9
XL	10	9,1	110	100,0
Celkem:	110	100,0	---	---

Zamyslete se nad interpretaci kumulativních četností (kumulativních relativních četností)!

Popisná statistika – Kvalitativní znak

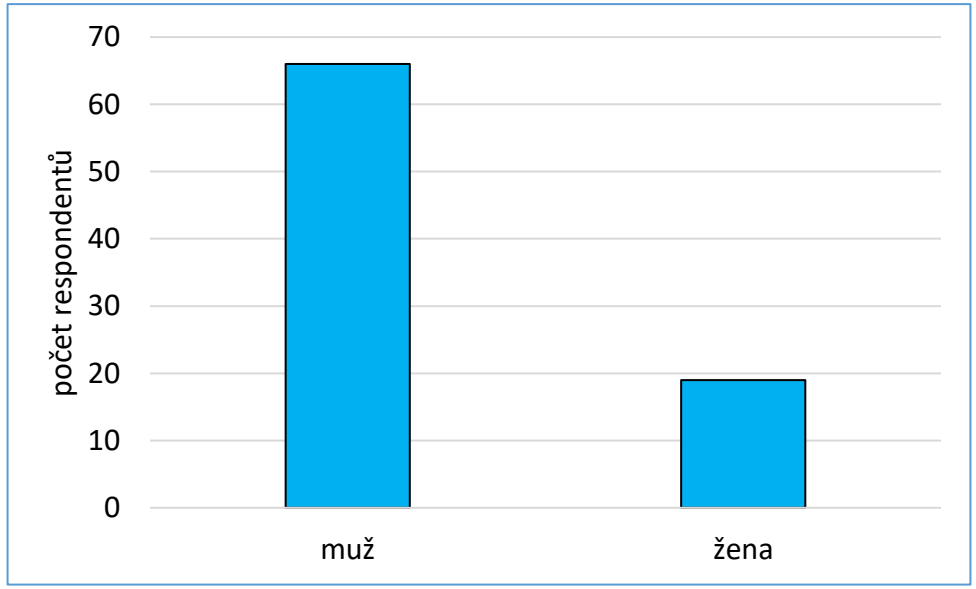
- **Tabulka četností**

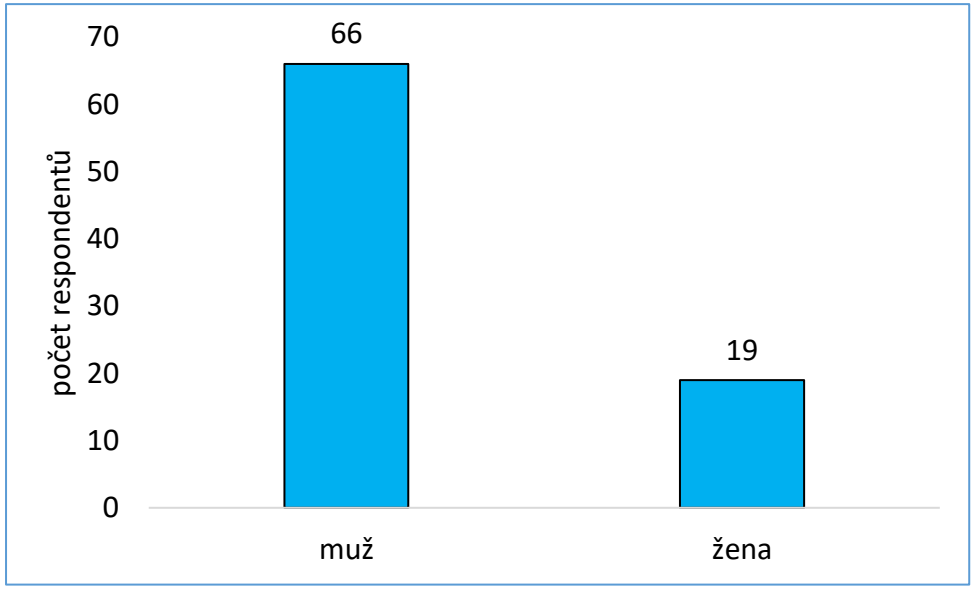
Tabulka četností		
Pohlaví	**Absolutní četnost**	**Relativní četnost (%)**
muž	66	78
žena	19	22
Celkem:	85	100

Jak data vizualizovat?

Popisná statistika – Kvalitativní znak

Sloupcový graf (Bar Chart)

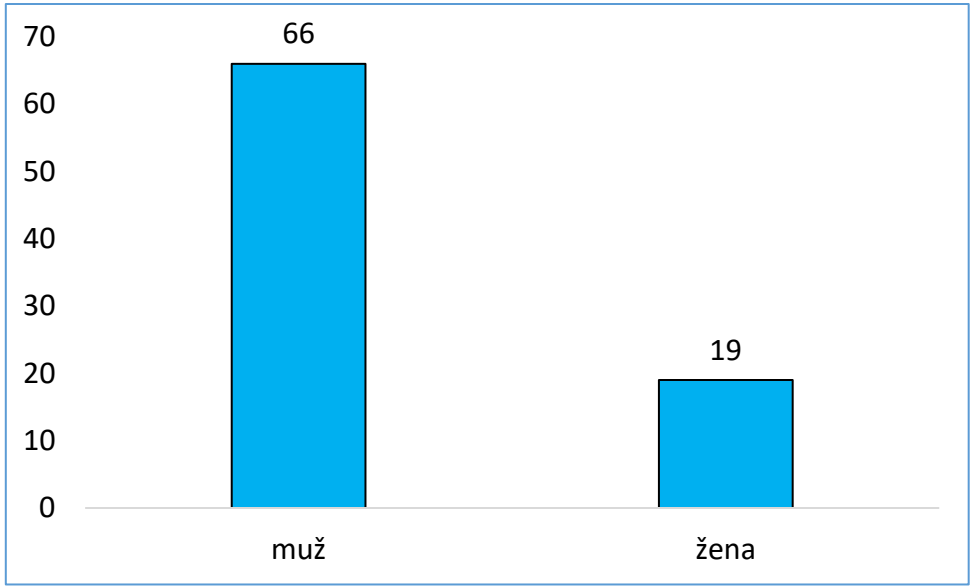


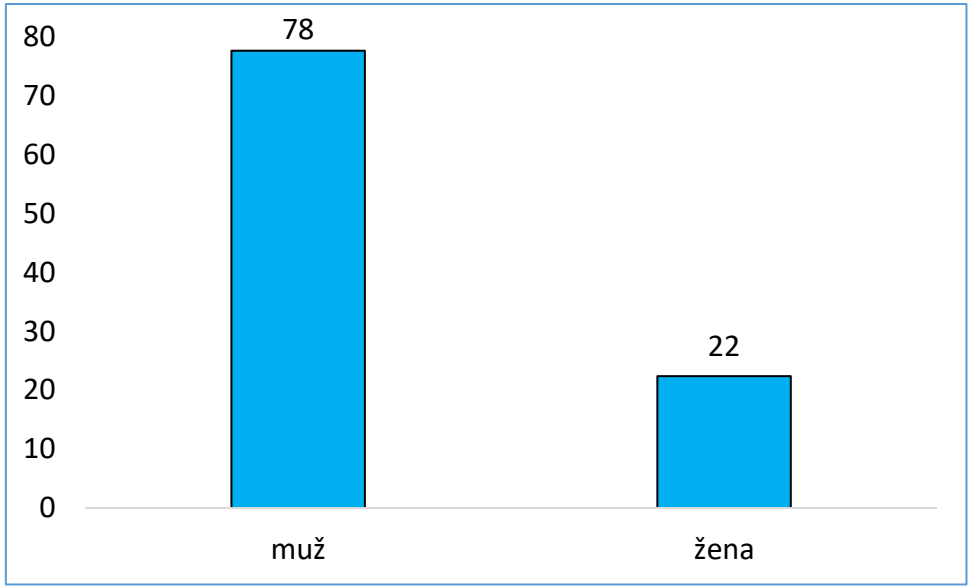


- Nejsou-li v grafu uvedeny absolutní (relativní) četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.

Popisná statistika – Kvalitativní znak

Sloupcový graf (Bar Chart)

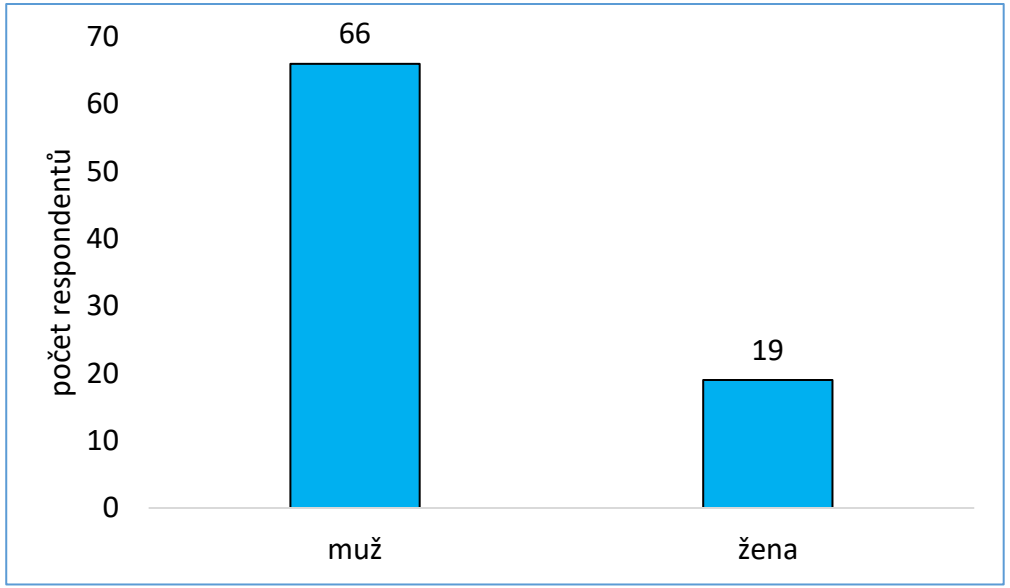


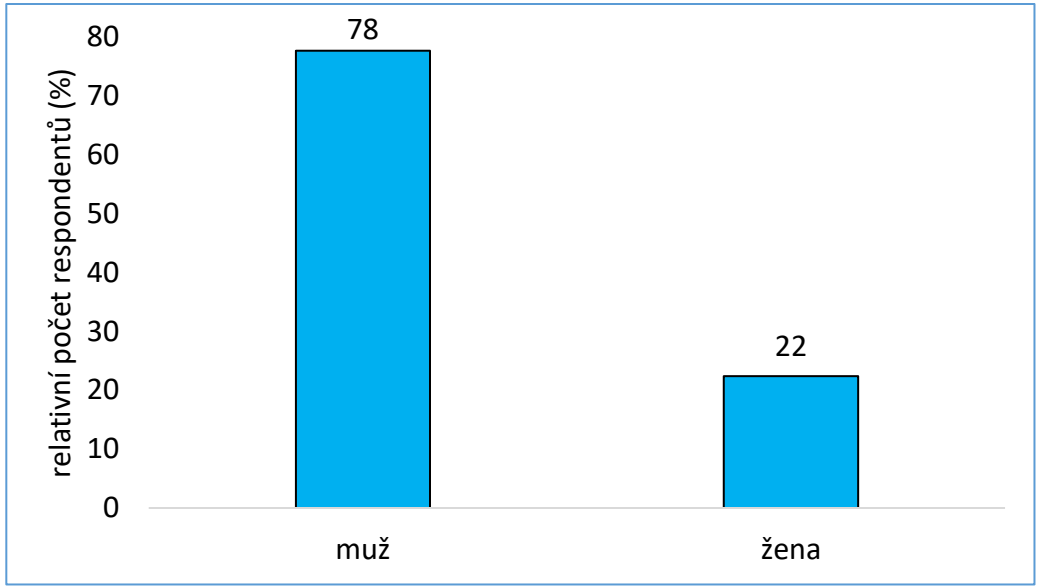


- Nejsou-li v grafu uvedeny absolutní (relativní) četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.
- Pozor na uvádění popisu os!

Popisná statistika – Kvalitativní znak

Sloupcový graf (Bar Chart)





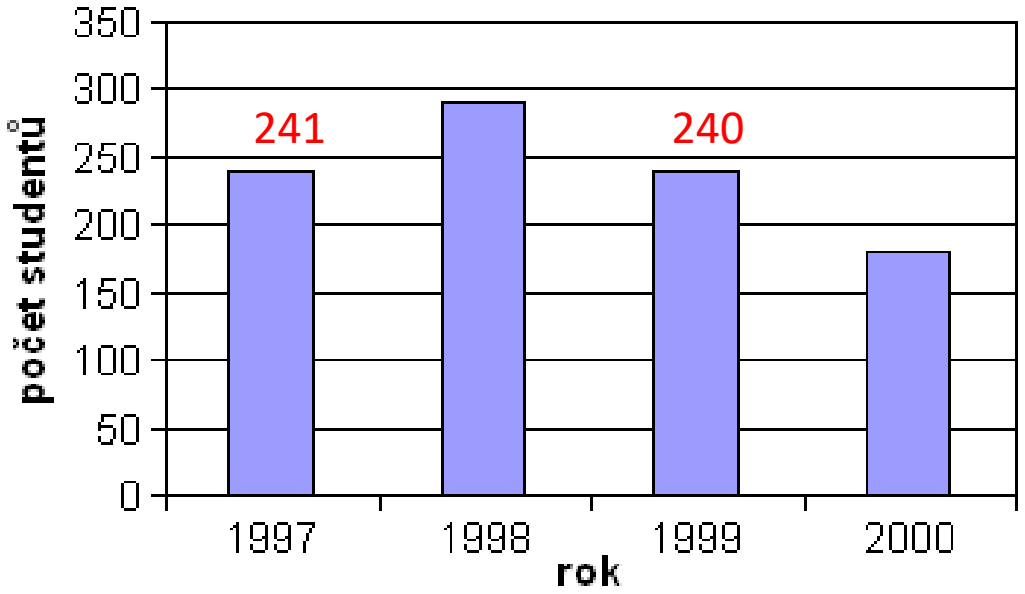
- Nejsou-li v grafu uvedeny absolutní (relativní) četnosti, obvykle je nedokážeme „od oka“ přesně odečíst.
- Pozor na uvádění popisu os!

Popisná statistika – Kvalitativní znak

Zdroj: Srovnávací testy pro žáky 9. tříd

Určete pravdivost tvrzení:

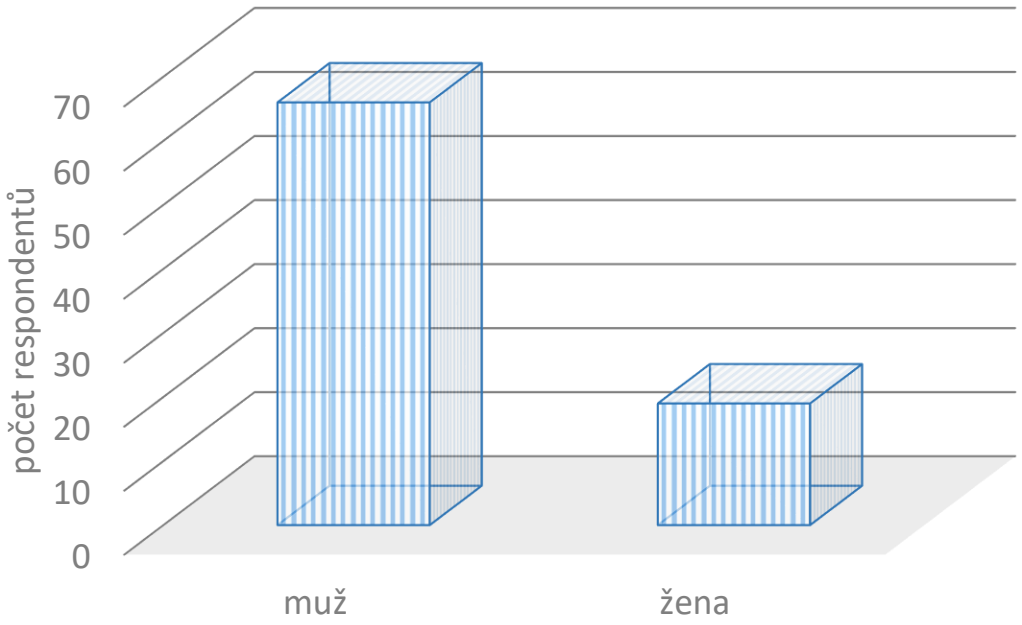
V žádných dvou letech nebyl počet studentů stejný.



Pozor na omezenou vypovídací schopnost grafů!

Popisná statistika – Kvalitativní znak

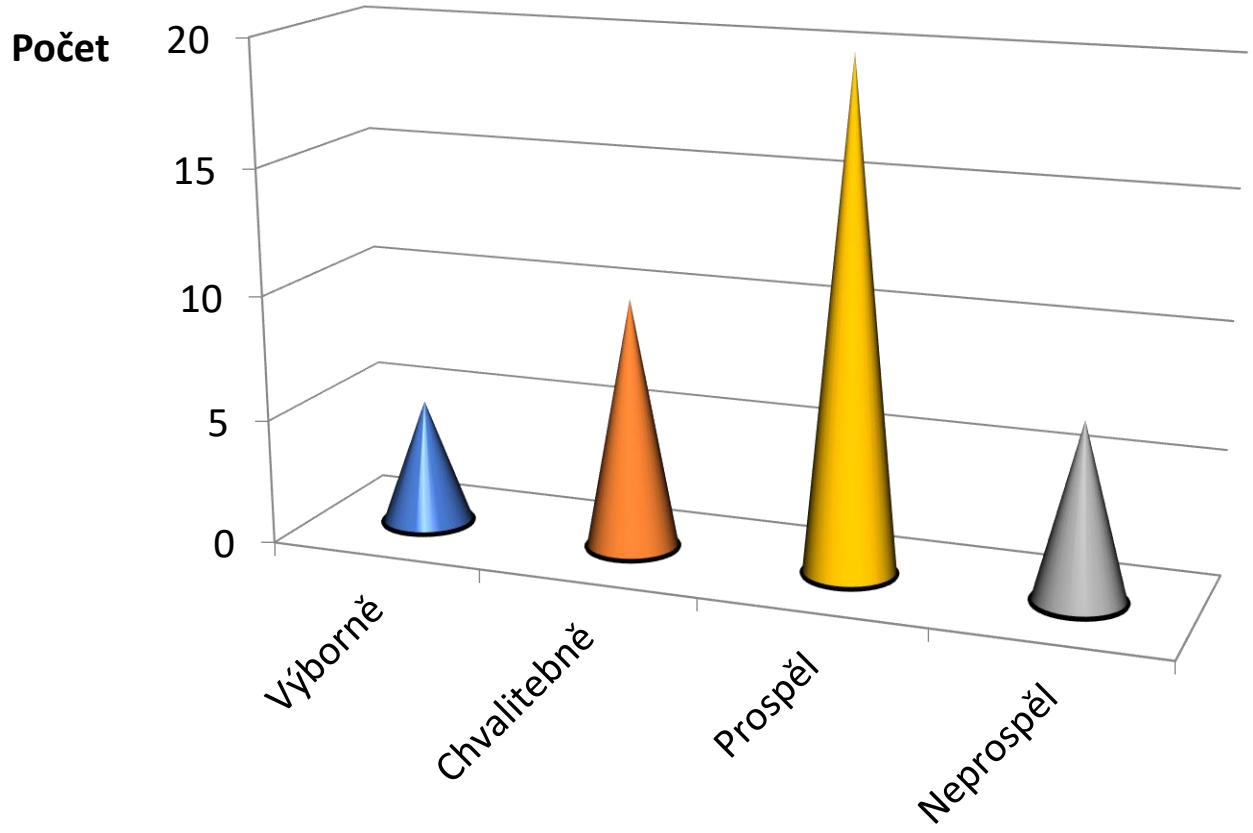
Sloupcový graf (Bar Chart)



Nezapomínejte, že méně mnohdy znamená více …

Popisná statistika – Kvalitativní znak

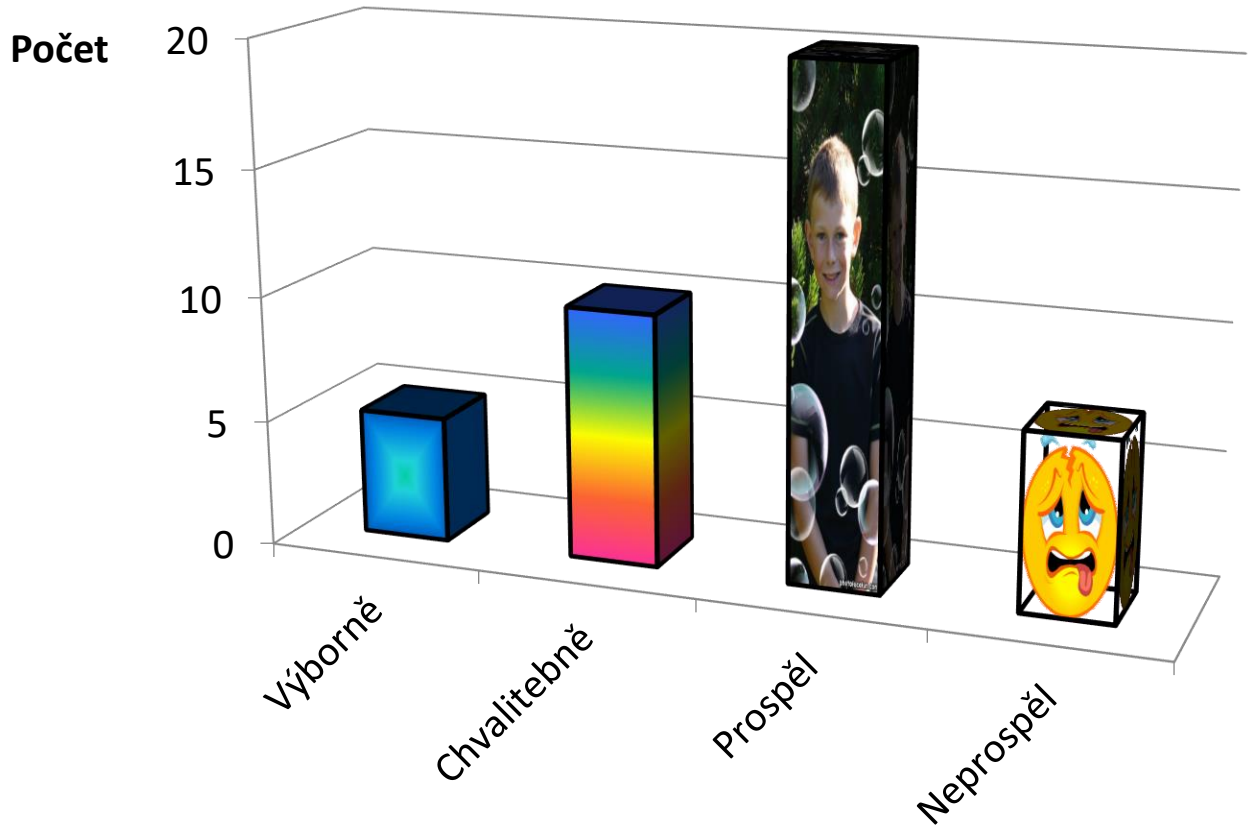
Sloupcový graf (Bar Chart)



I takto může vypadat sloupcový graf …

Popisná statistika – Kvalitativní znak

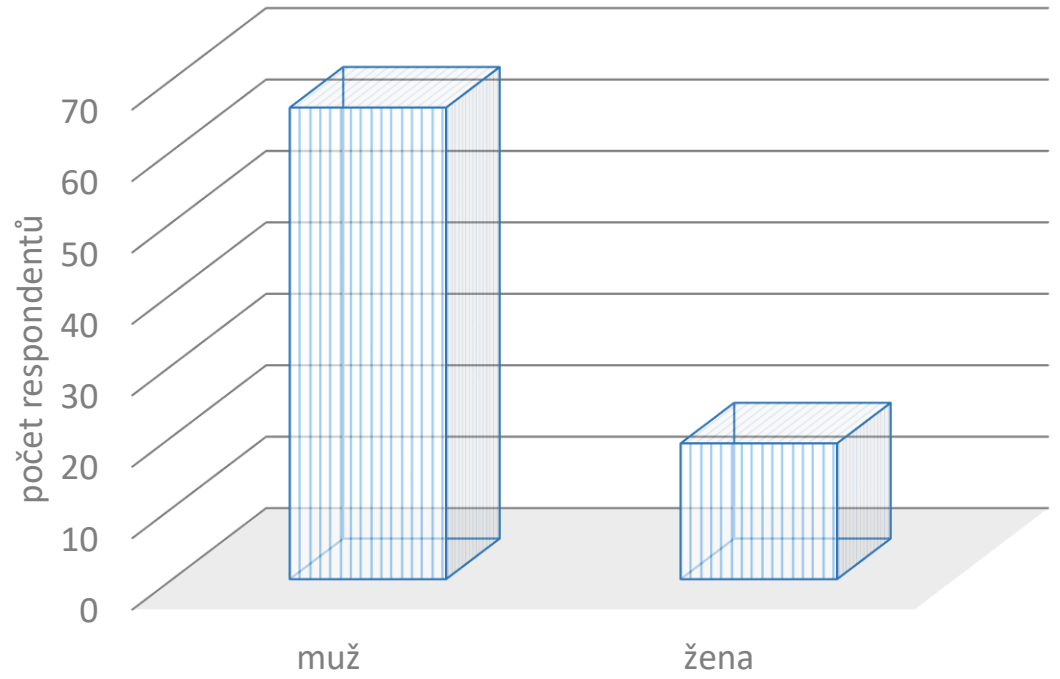
Sloupcový graf (Bar Chart)



I takto může vypadat sloupcový graf …

Popisná statistika – Kvalitativní znak

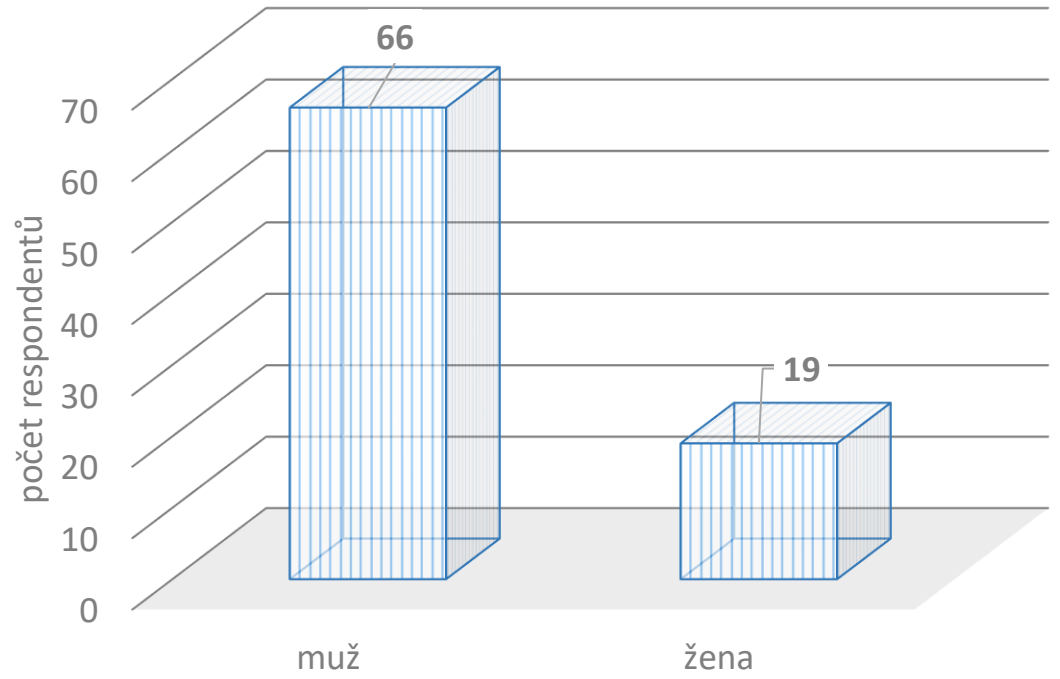
Sloupcový graf (Bar Chart)



Kolik bylo mezi respondenty mužů, resp. žen?

Popisná statistika – Kvalitativní znak

Sloupcový graf (Bar Chart)

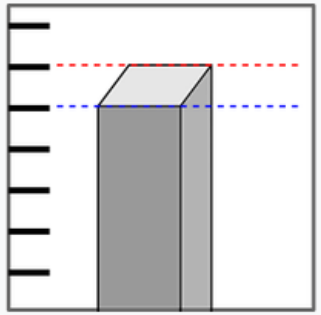


Kolik bylo mezi respondenty mužů, resp. žen?

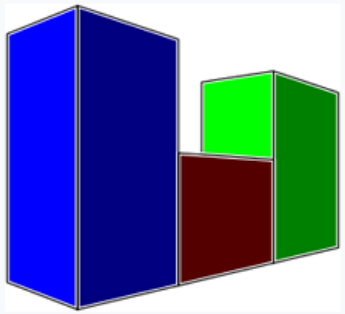
Popisná statistika – Kvalitativní znak

3D sloupcový graf

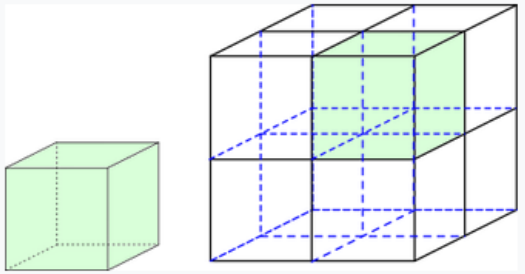
Zdroj: Whitbread, David (2001). The design manual (2nd ed.). Sydney: University of New South Wales Press. ISBN 0868406589.



The third dimension may confuse readers[9]



The blue bar in the front appears larger than the green bar in the back due to perspective, despite having the same value

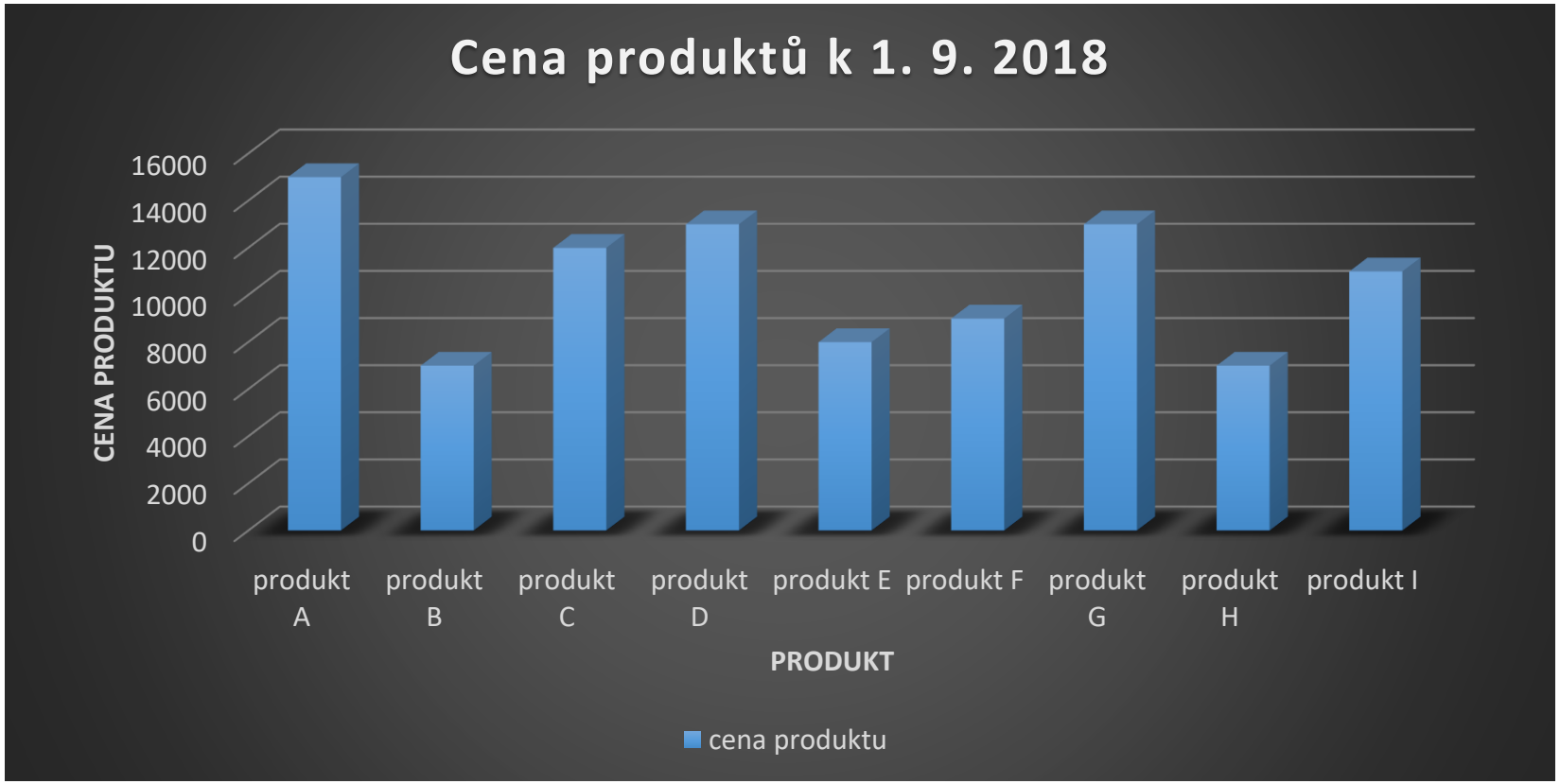


When scaling in three dimensions, the effect of the change is cubed

Opravdu musí být v každé efektní / efektivní prezentaci?

Popisná statistika – Kvalitativní znak

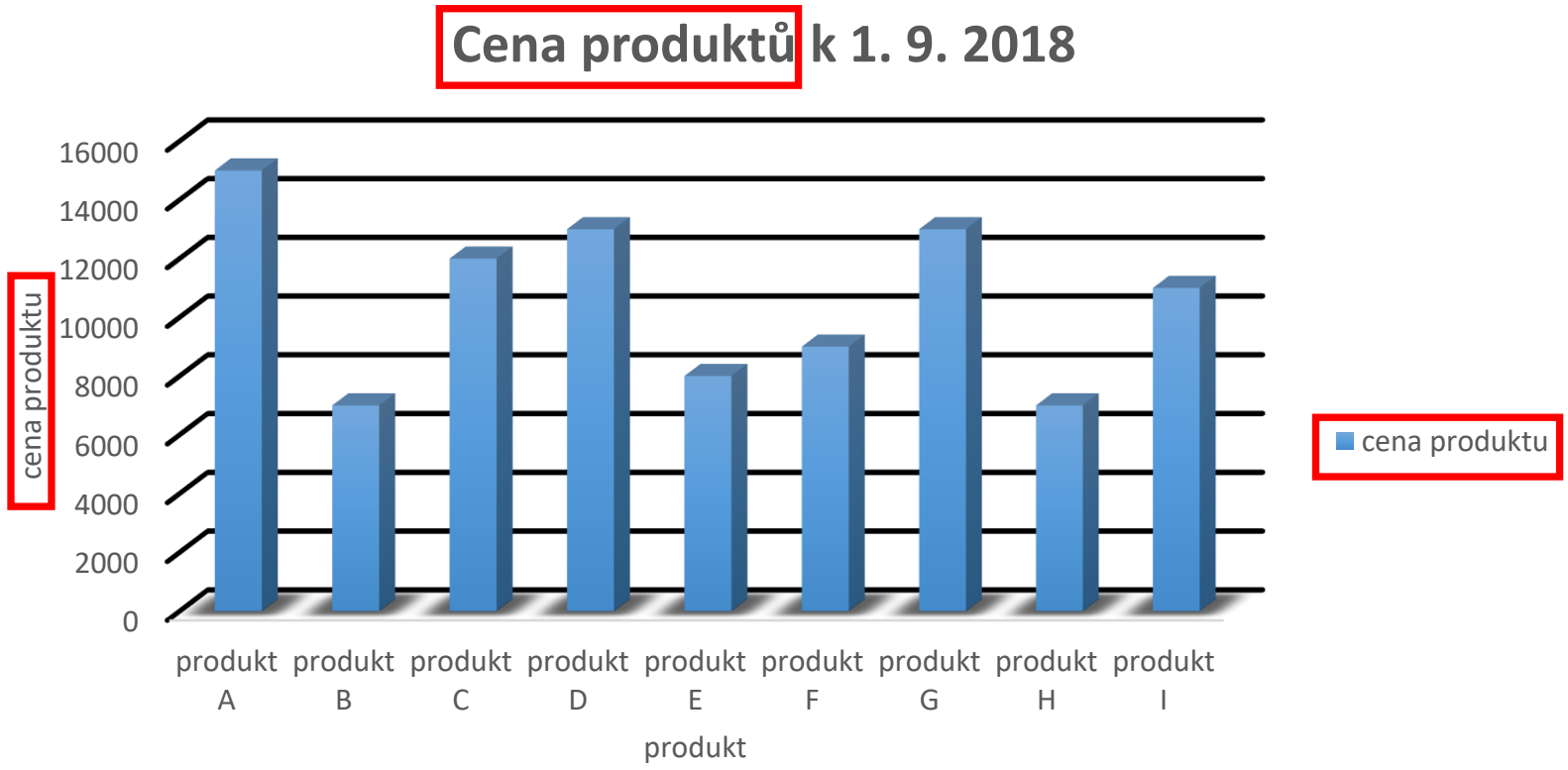
Tipy nejen pro sloupcové grafy…



Nepoužívejte barevné pozadí grafů!

Popisná statistika – Kvalitativní znak

Tipy nejen pro sloupcové grafy…

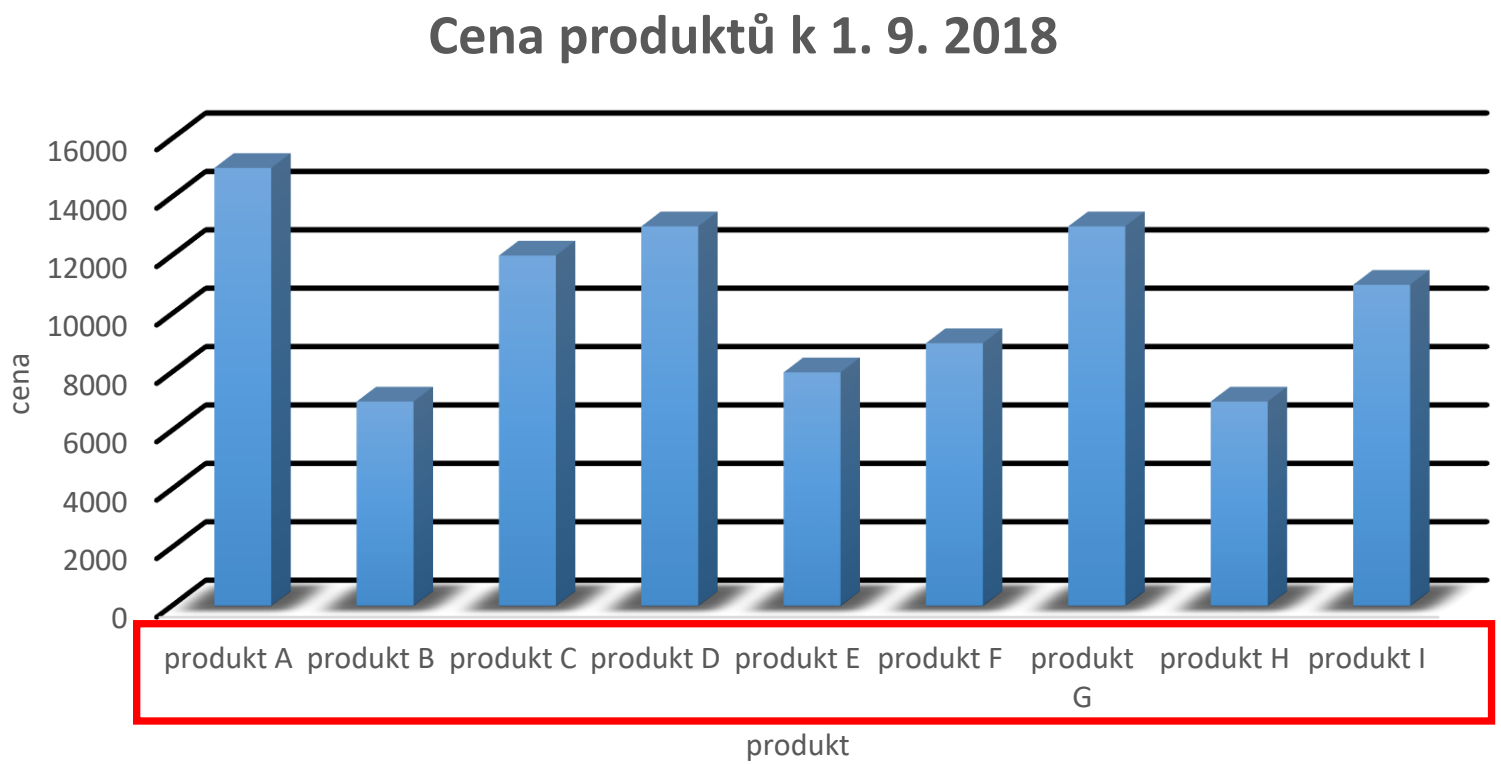


Neopakujte informace!

V tomto případě je legenda zcela nadbytečná. V případě, že ji potřebujete, zvažte její umístění!

Popisná statistika – Kvalitativní znak

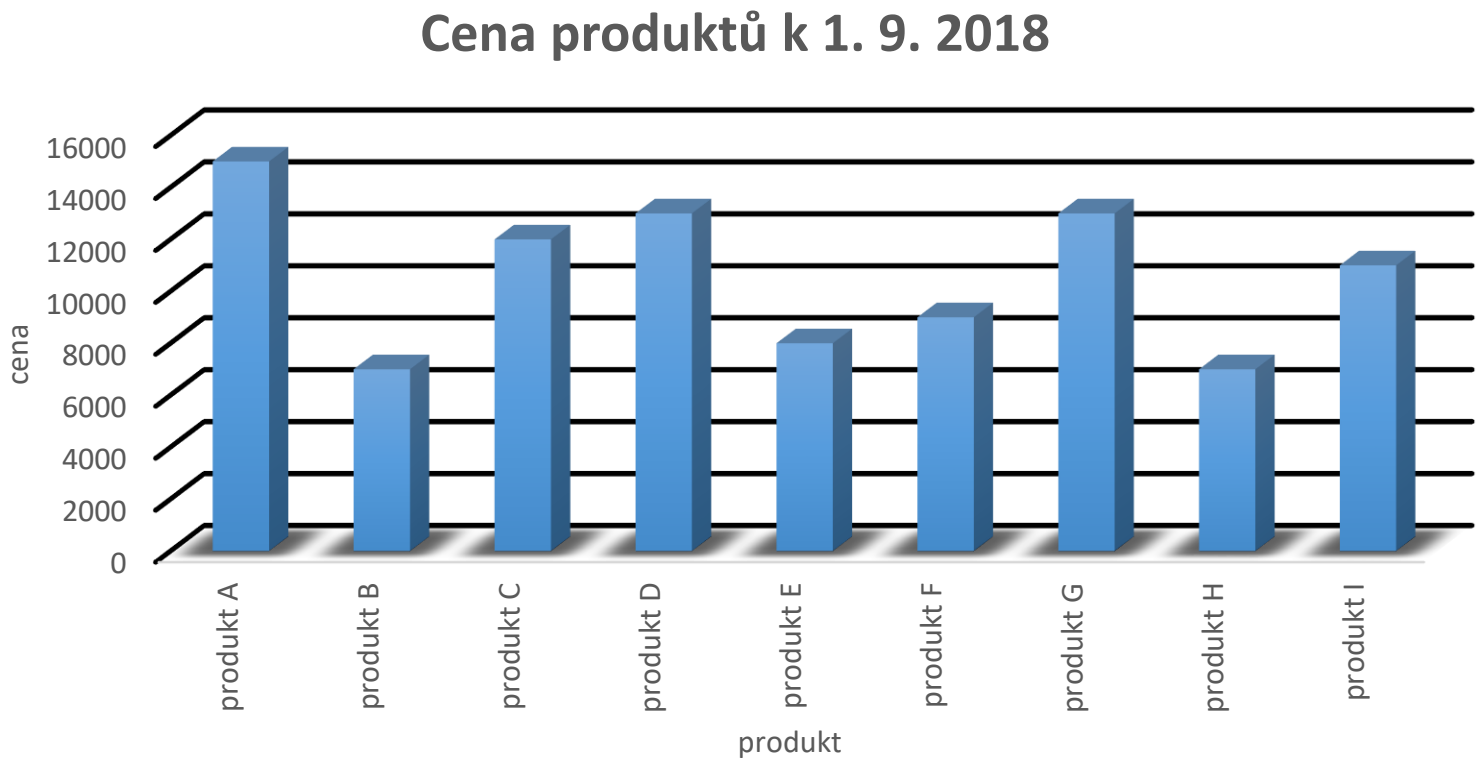
Tipy nejen pro sloupcové grafy…



Zvažte úpravu popisku horizontální osy!

Popisná statistika – Kvalitativní znak

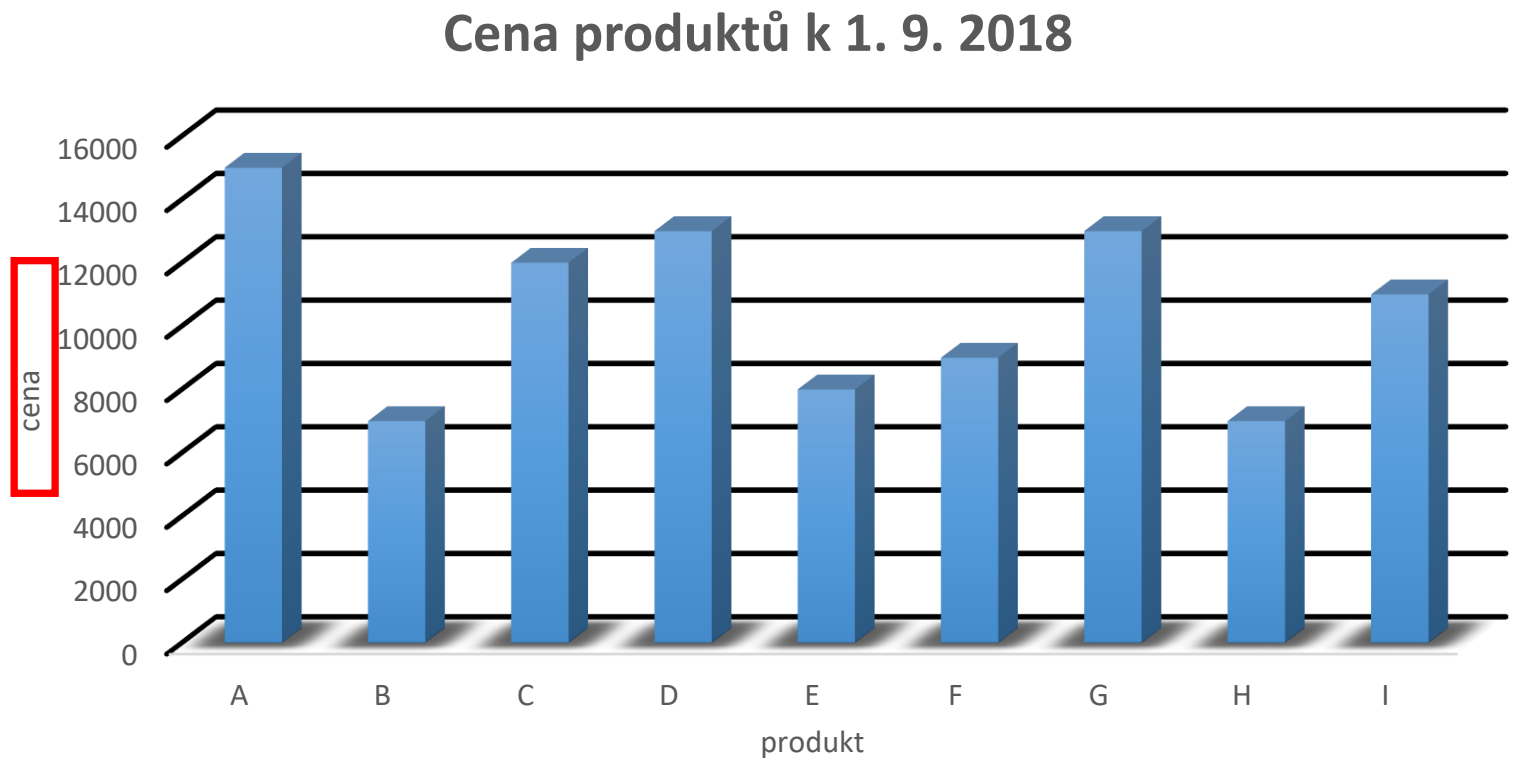
Tipy nejen pro sloupcové grafy…



Zvažte úpravu popisku horizontální osy!

Popisná statistika – Kvalitativní znak

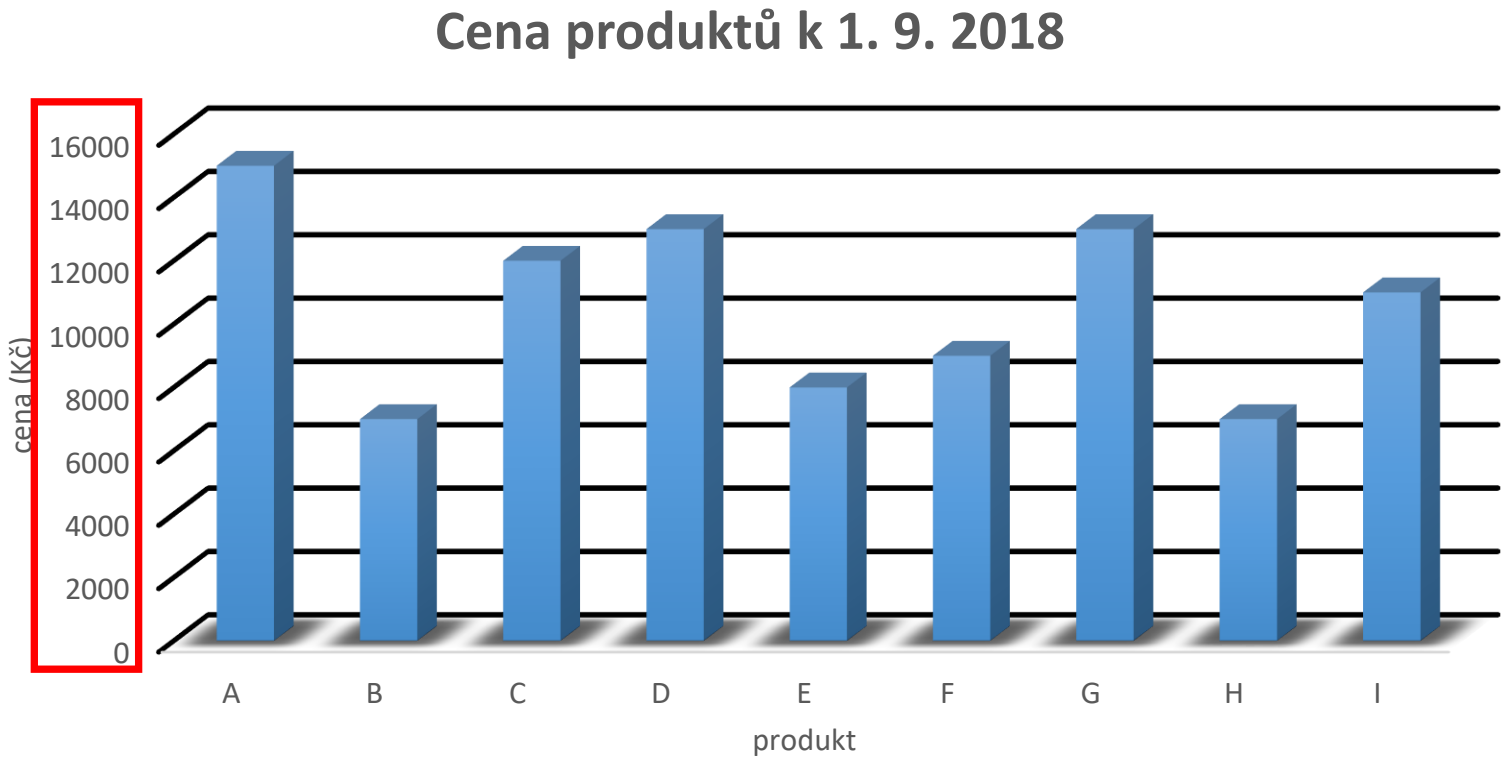
Tipy nejen pro sloupcové grafy…



Vždy uvádějte jednotky!

Popisná statistika – Kvalitativní znak

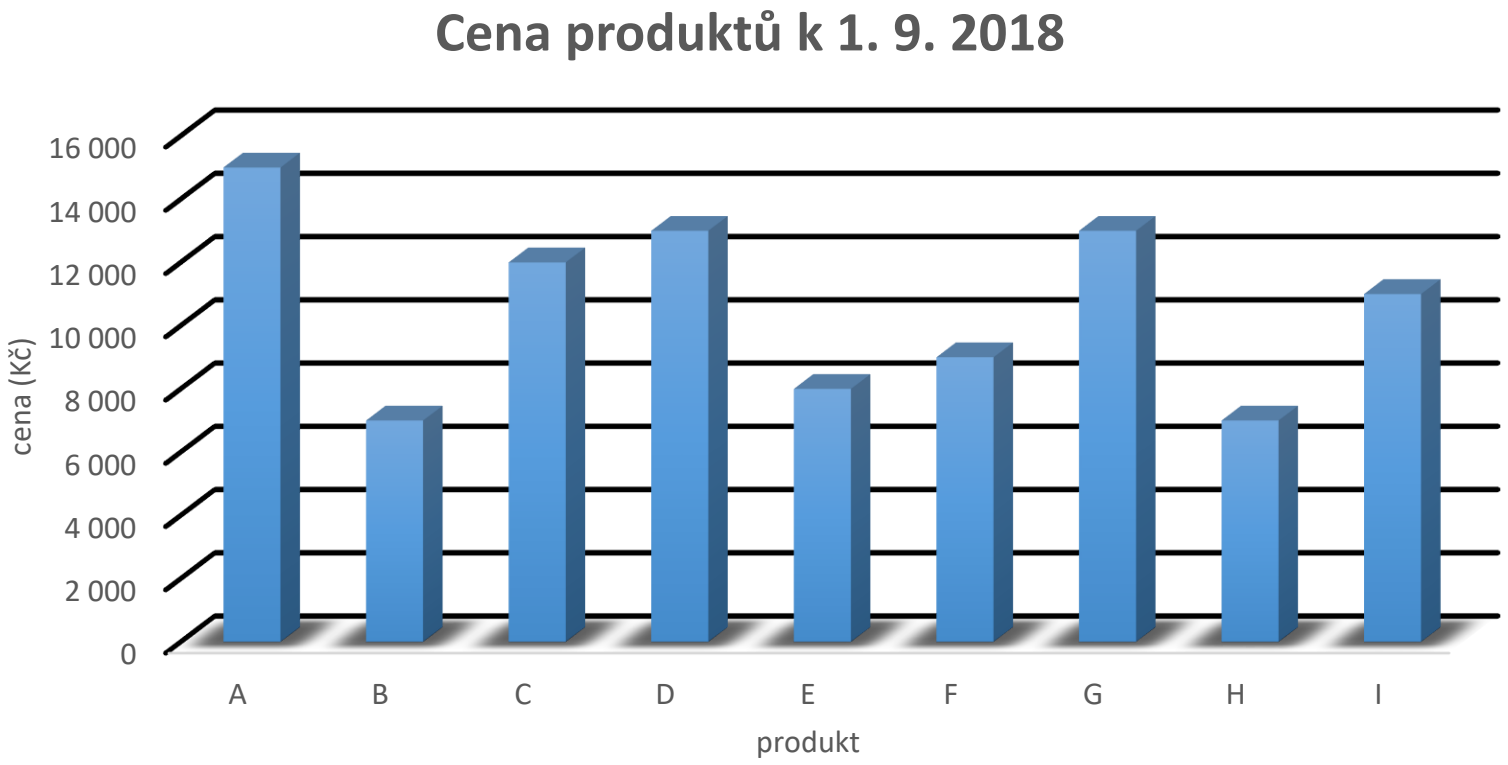
Tipy nejen pro sloupcové grafy…



Nepoužívejte neefektivní nuly! (Nebo alespoň použijte oddělovače tisíců!)

Popisná statistika – Kvalitativní znak

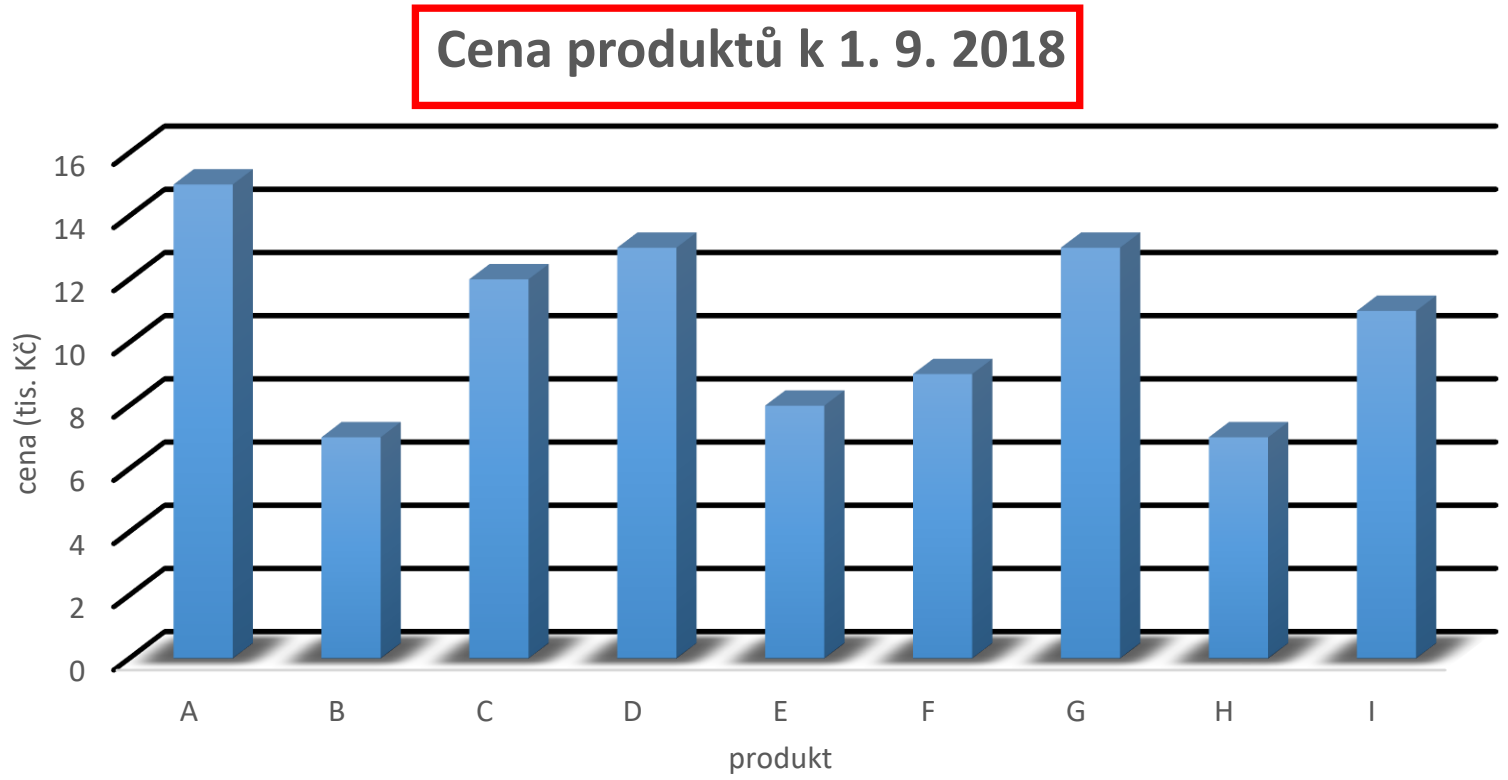
Tipy nejen pro sloupcové grafy…



Nepoužívejte neefektivní nuly! (Nebo alespoň použijte oddělovače tisíců!)

Popisná statistika – Kvalitativní znak

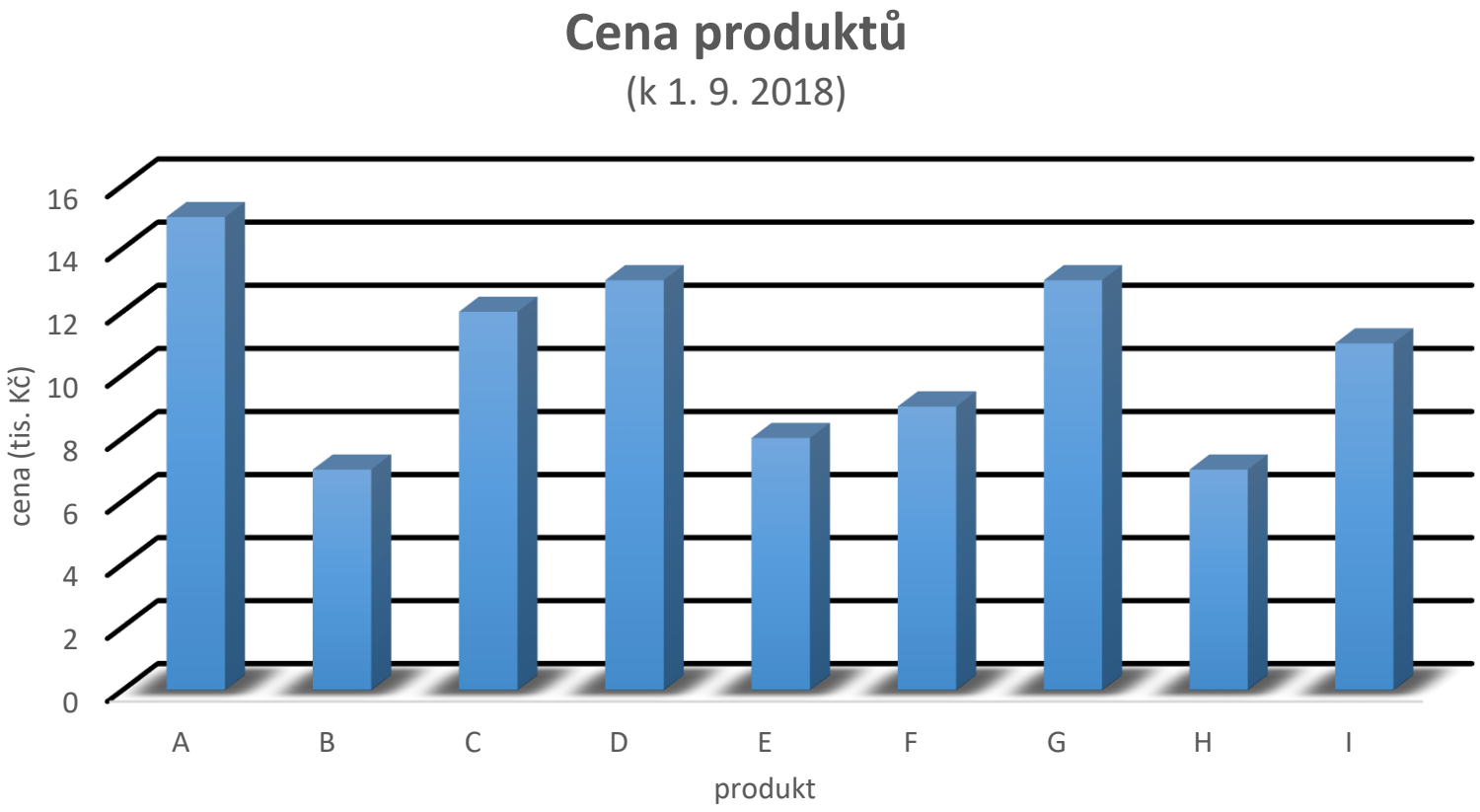
Tipy nejen pro sloupcové grafy…



Používejte strukturované nadpisy!

Popisná statistika – Kvalitativní znak

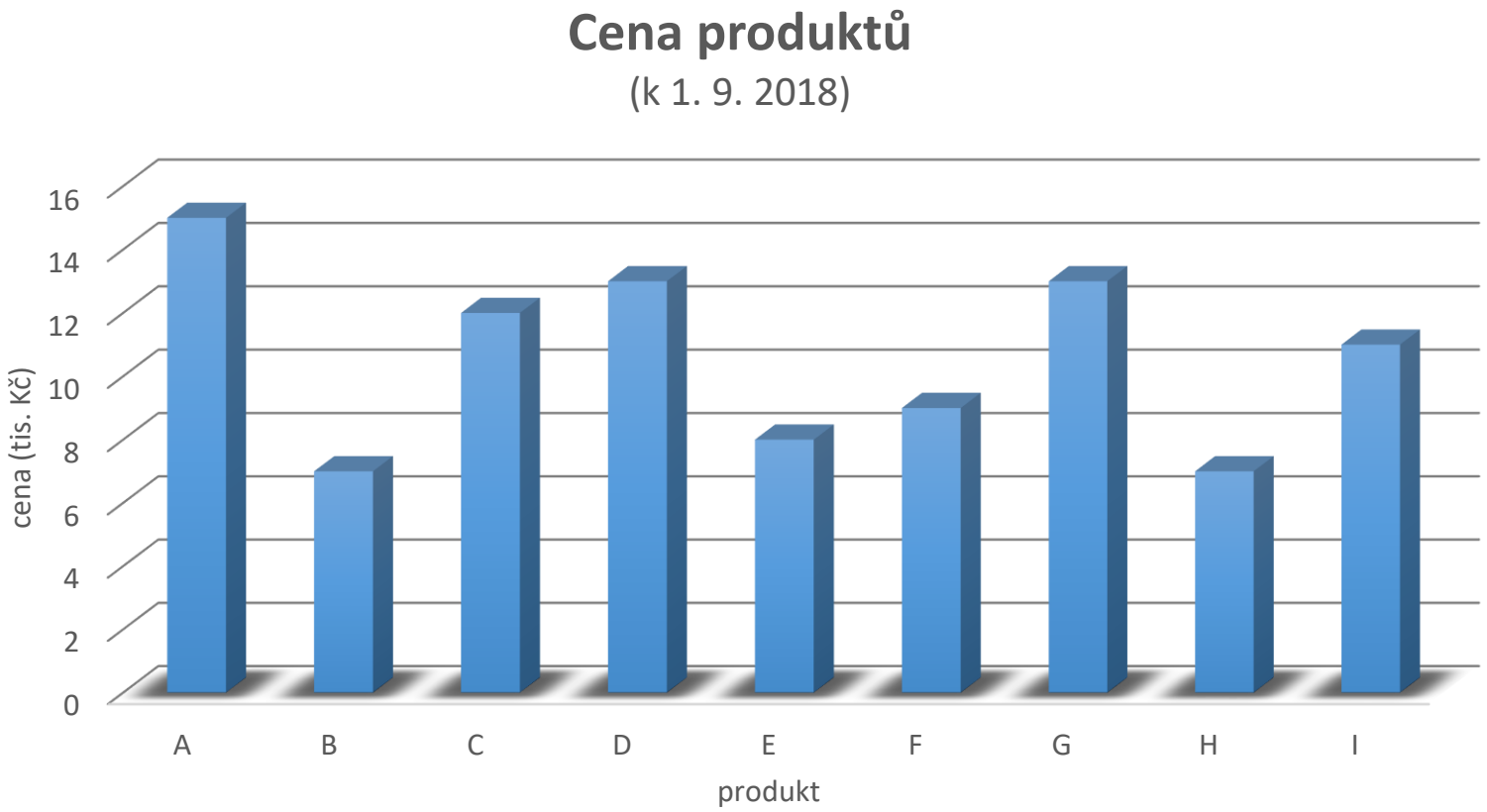
Tipy nejen pro sloupcové grafy…



Zvažte zvýraznění mřížky!

Popisná statistika – Kvalitativní znak

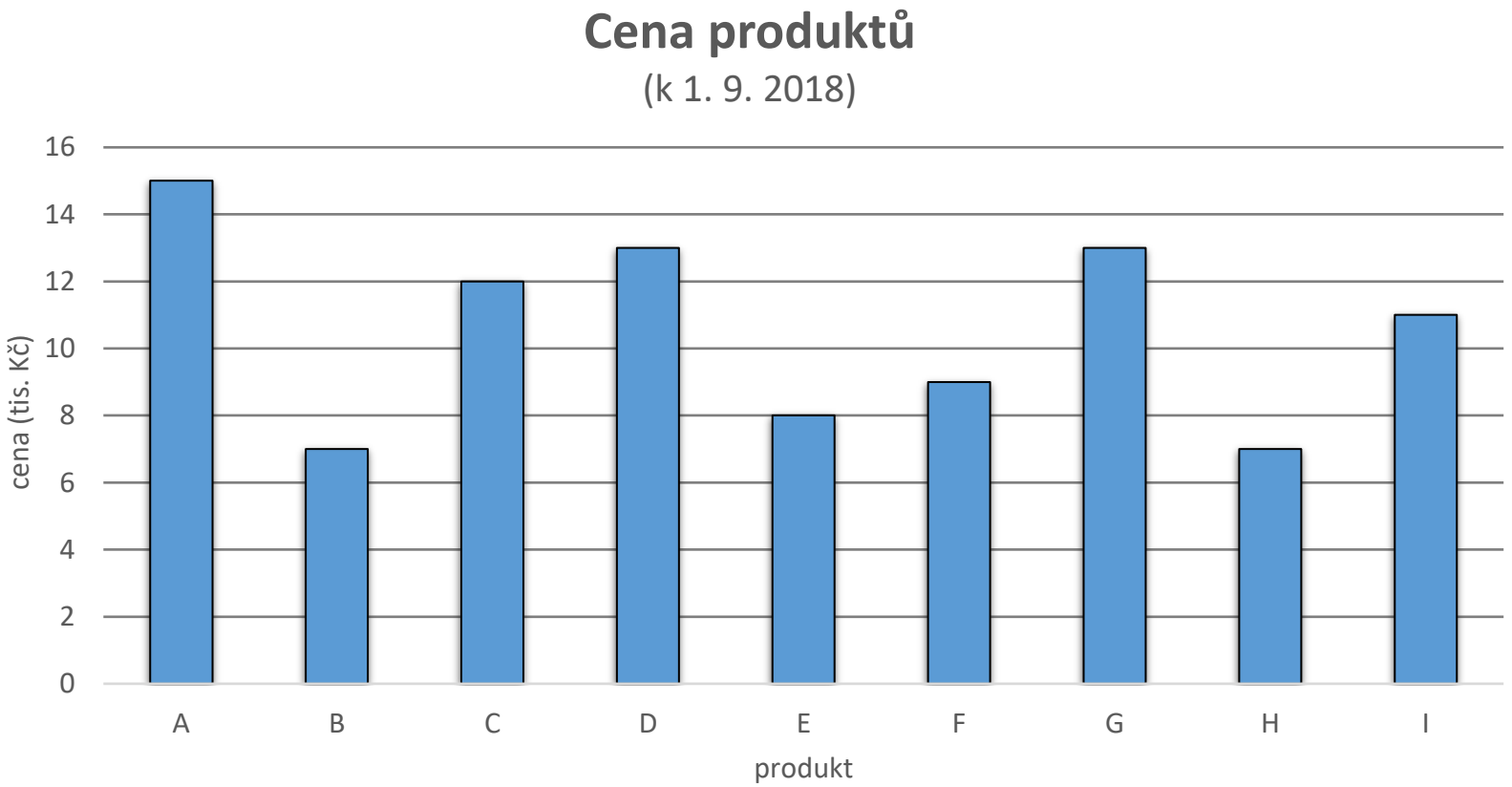
Tipy nejen pro sloupcové grafy…



Nepoužívejte 3D grafy!

Popisná statistika – Kvalitativní znak

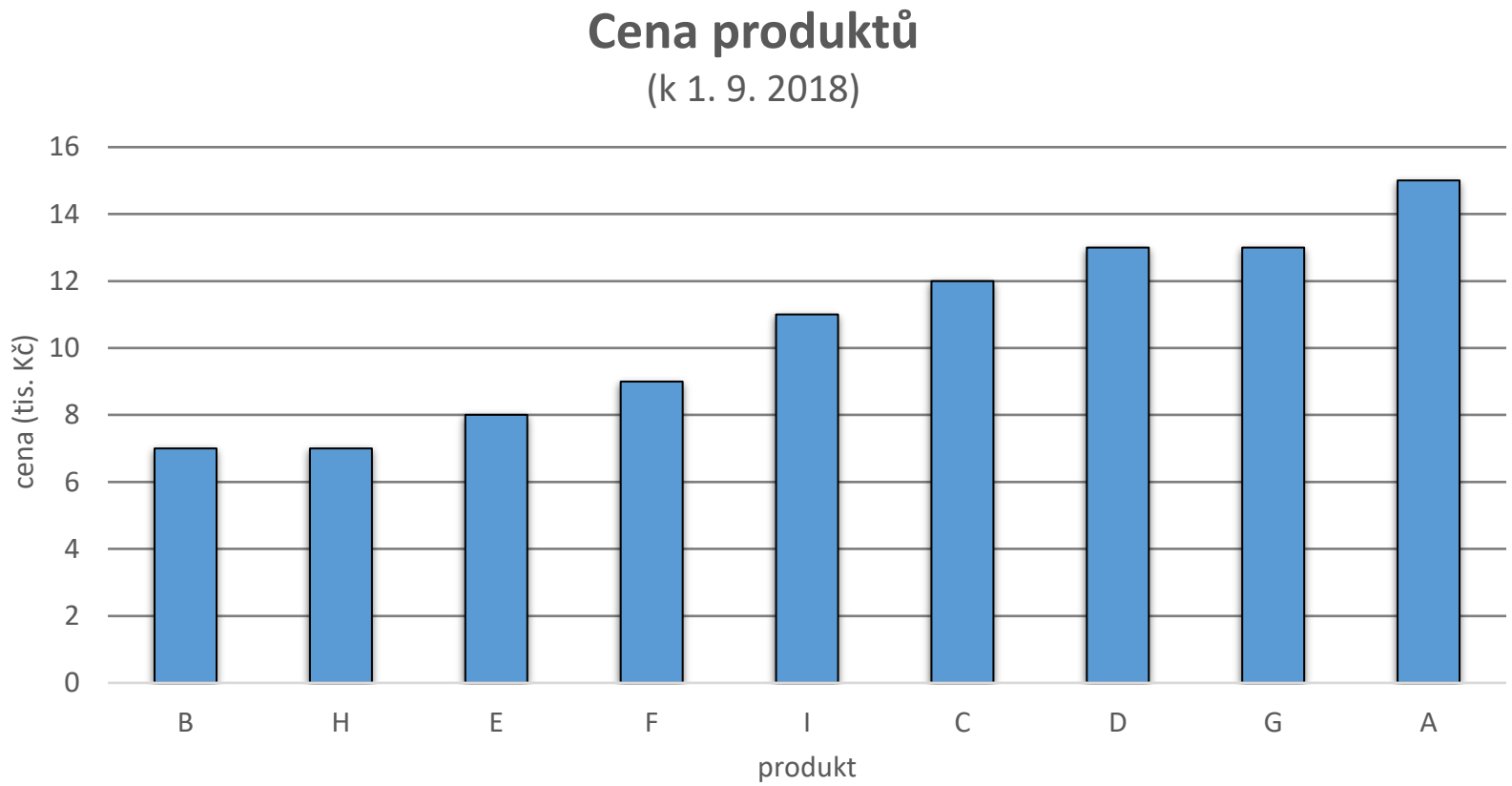
Tipy nejen pro sloupcové grafy…



Vždy přemýšlejte nad tím, co chcete na grafu ukázat!!!

Popisná statistika – Kvalitativní znak

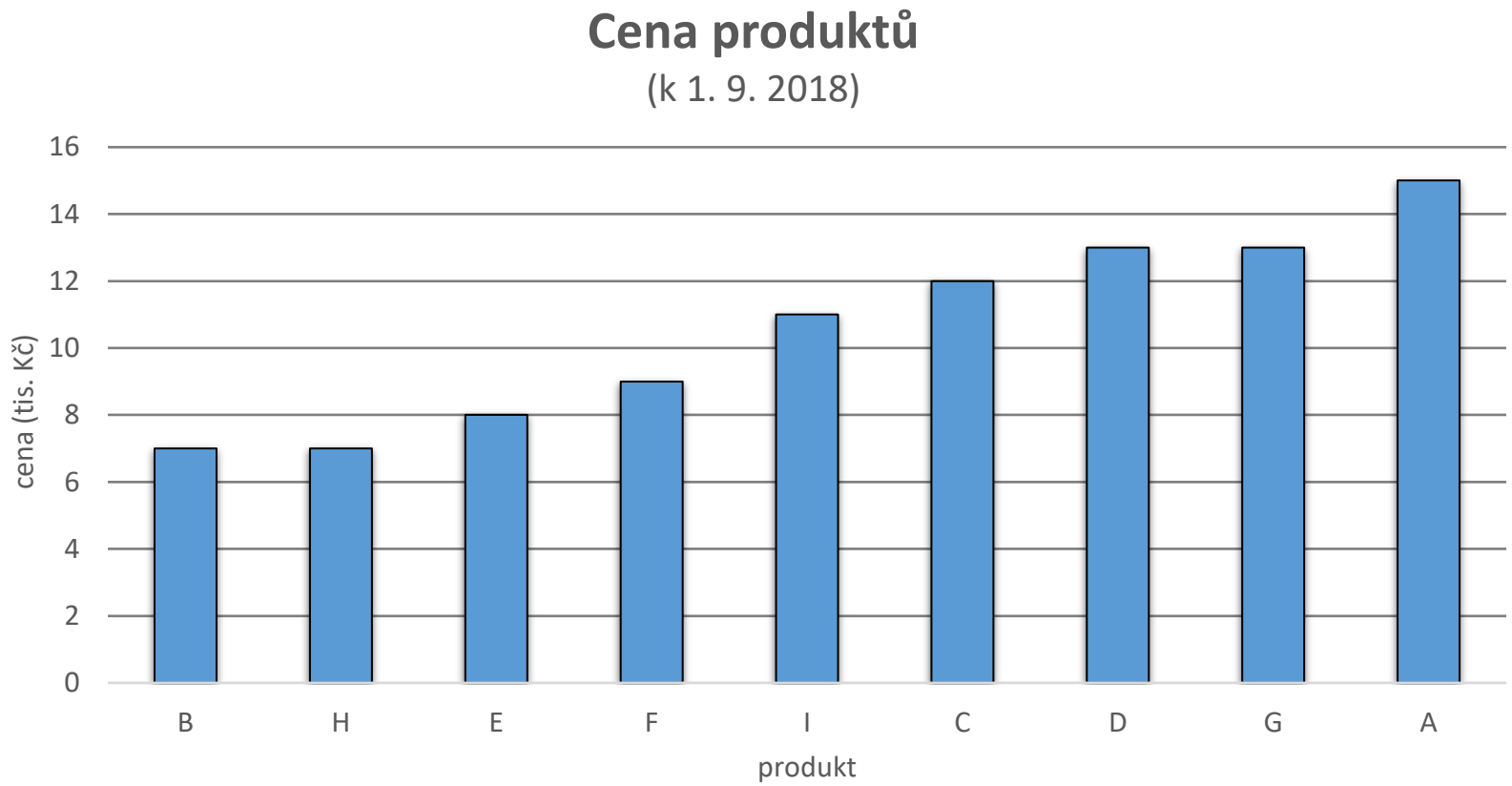
Tipy nejen pro sloupcové grafy…



Varianty nominálního znaku je vhodné řadit dle četností,
varianty ordinálního znaku dle přirozeného uspořádání!

Popisná statistika – Kvalitativní znak

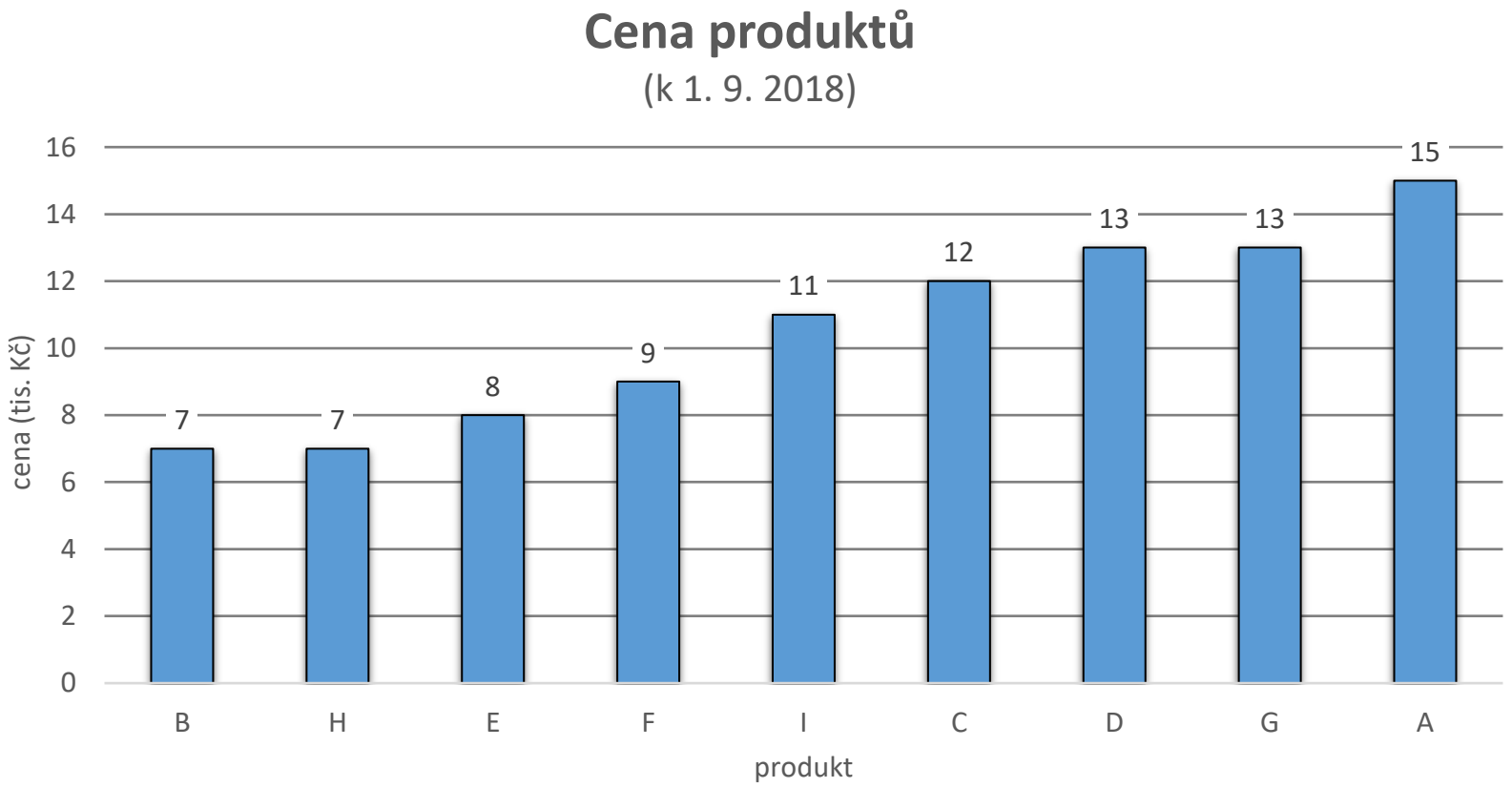
Tipy nejen pro sloupcové grafy…



Nejsou-li v grafu uvedeny absolutní četnosti, obvykle je nedokážeme „od oka“ přesně odečíst!

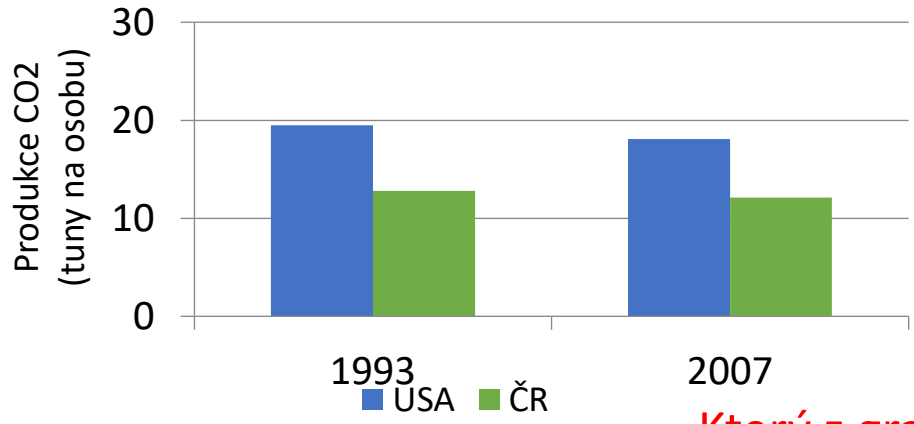
Popisná statistika – Kvalitativní znak

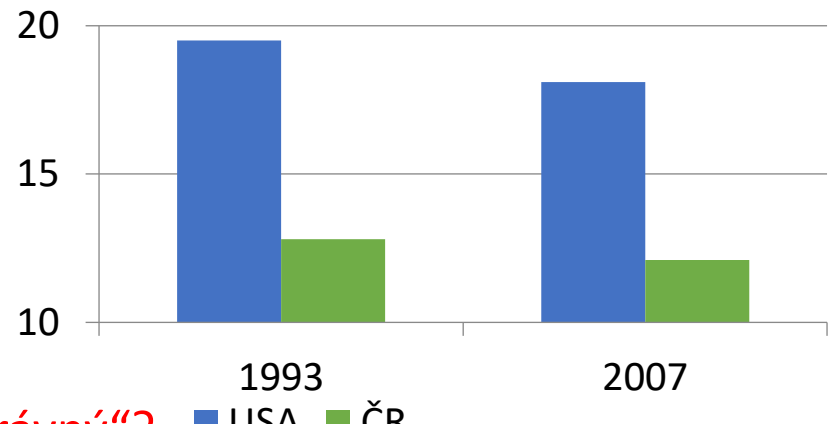
Tipy nejen pro sloupcové grafy…



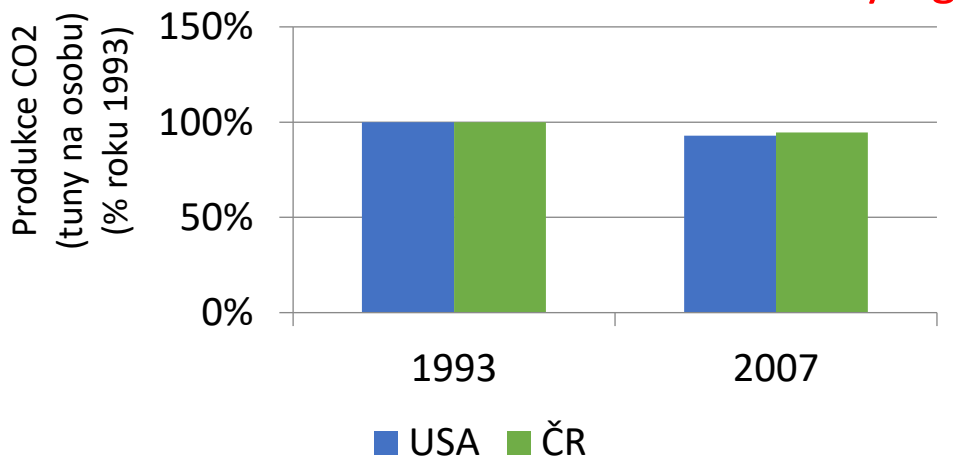
Popisná statistika – Kvalitativní znak

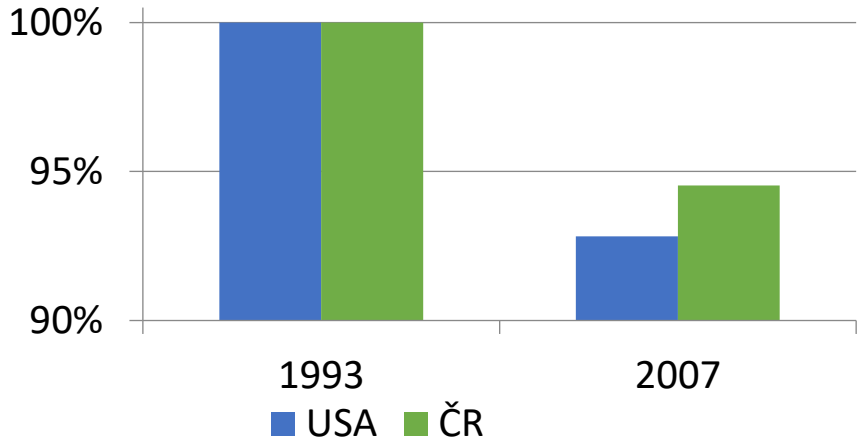
Tipy nejen pro sloupcové grafy…





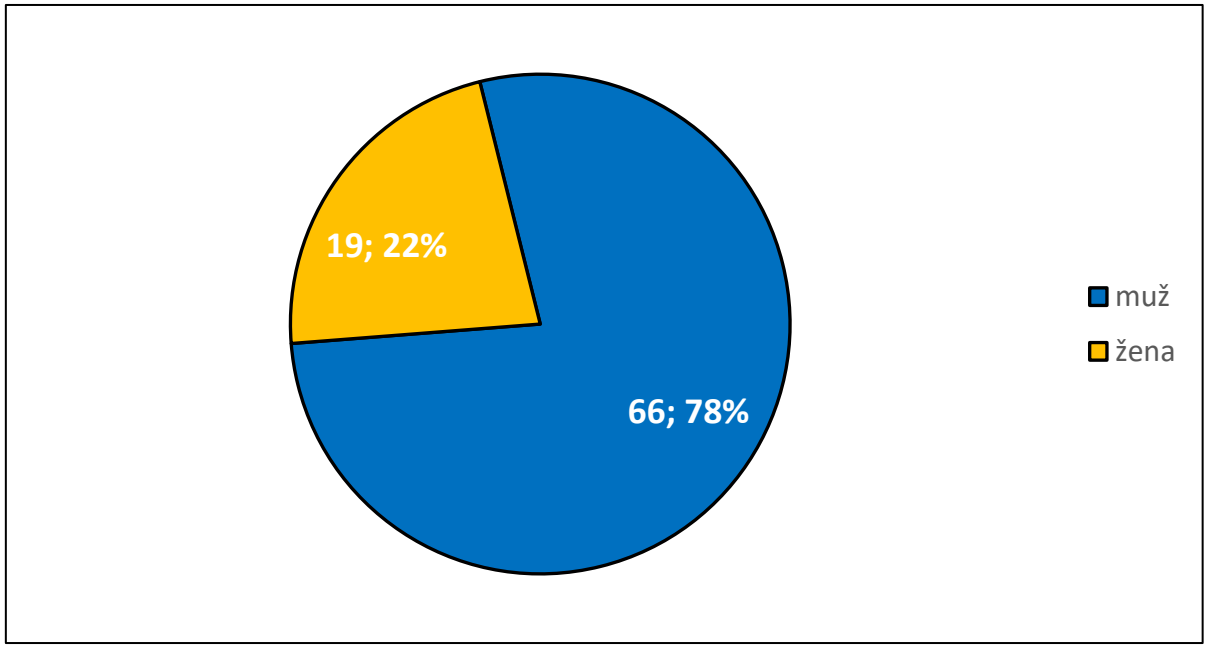
Který z grafů je „správný“?





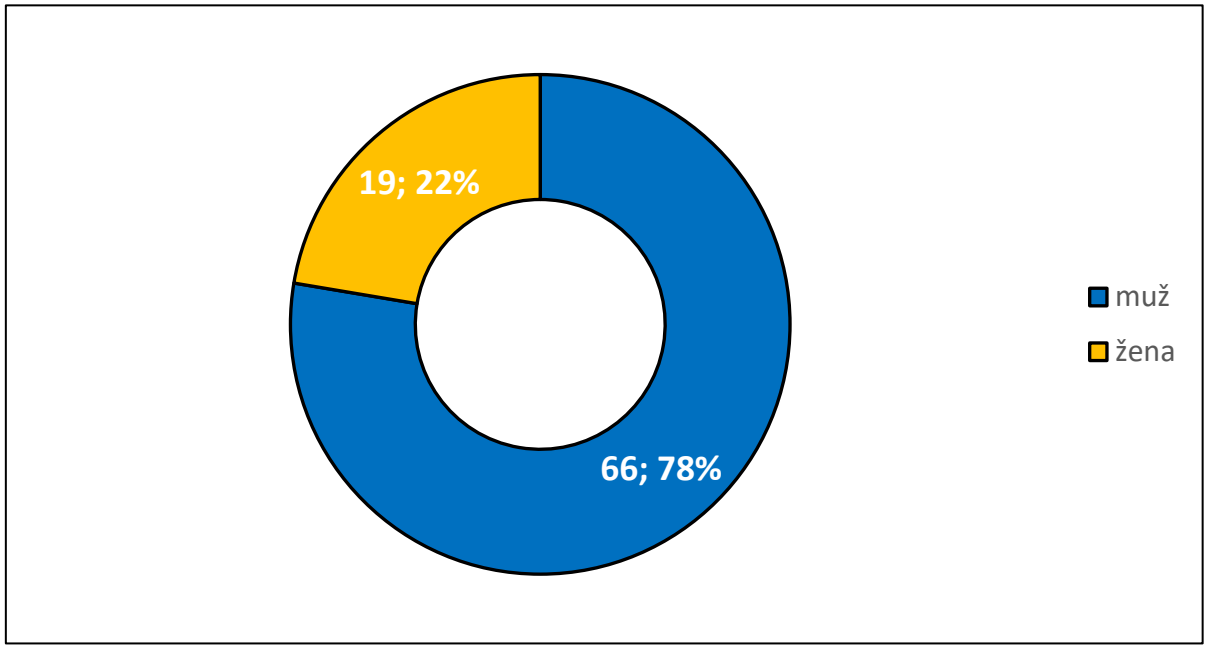
Popisná statistika – Kvalitativní znak

Výsečový graf



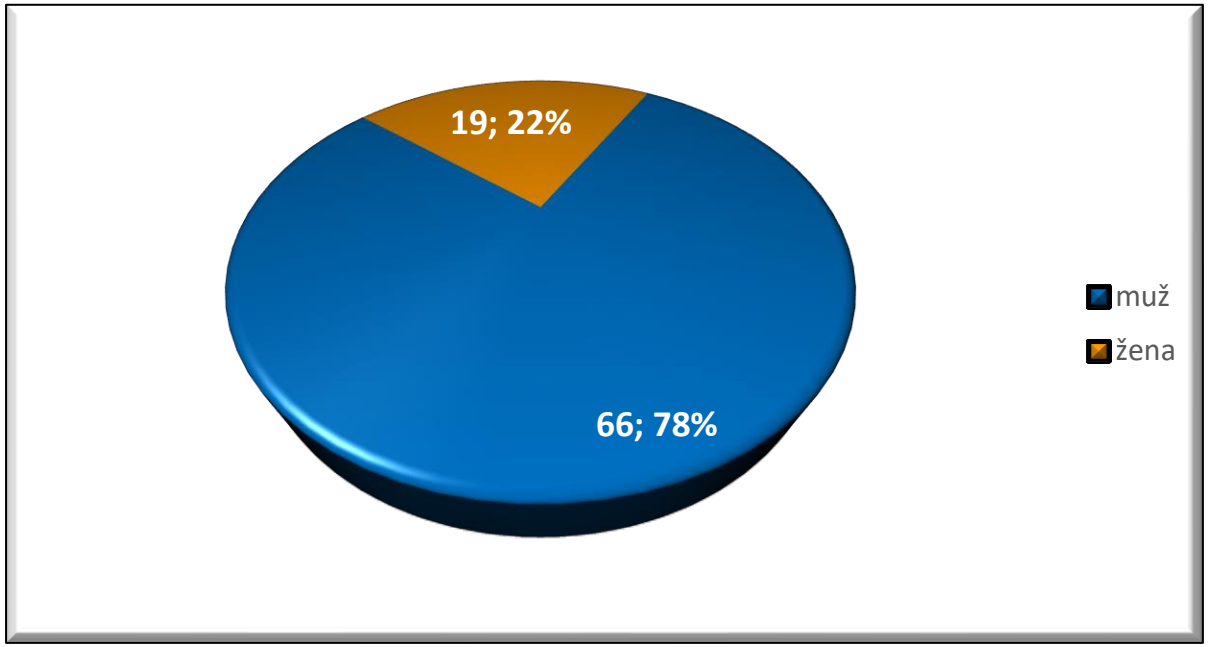
Popisná statistika – Kvalitativní znak

Prstencový graf



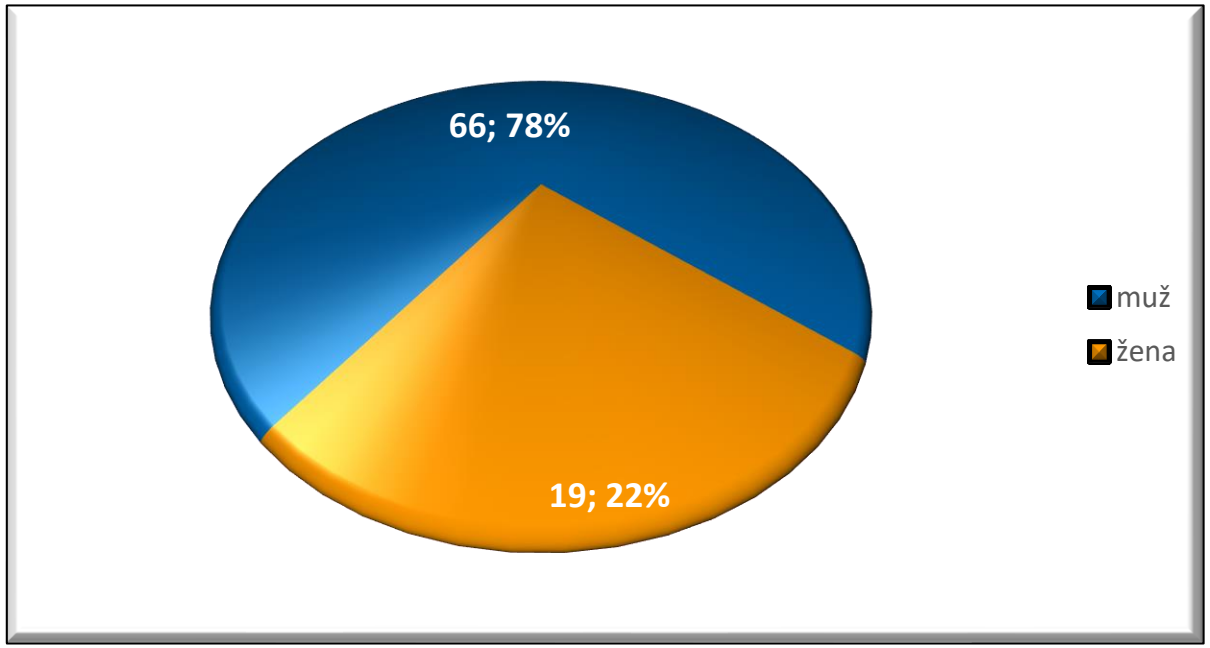
Popisná statistika – Kvalitativní znak

3D výsečový graf



Popisná statistika – Kvalitativní znak

3D výsečový graf



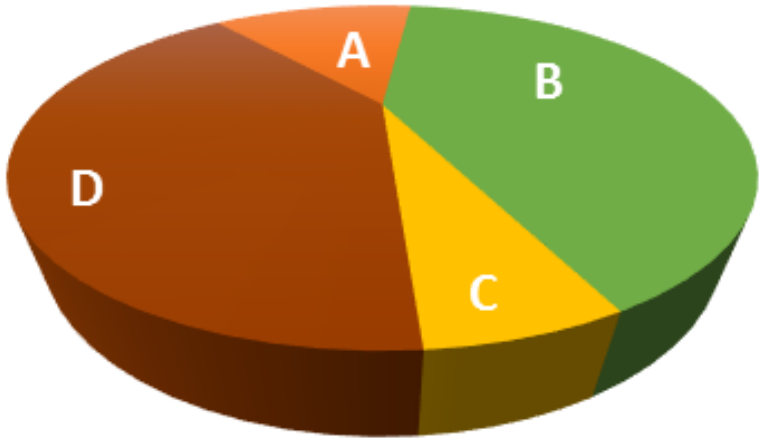
Pozor na vypovídací schopnost 3D grafů!

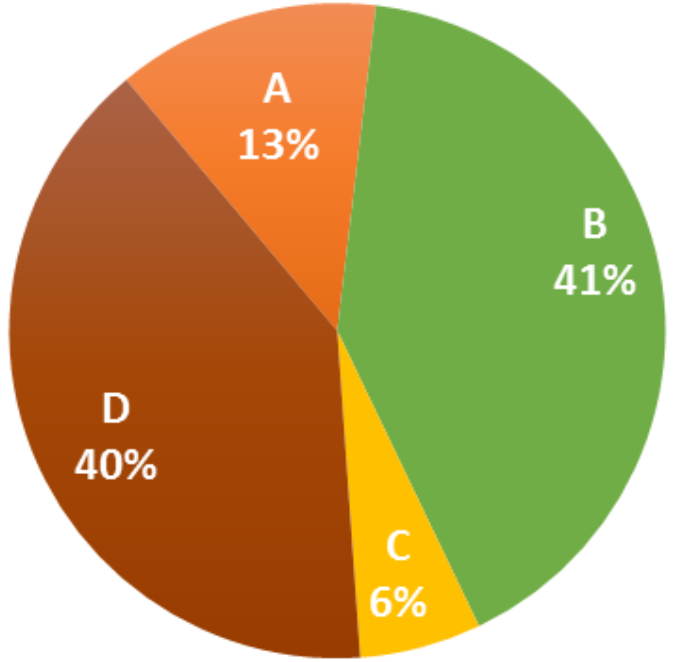
Popisná statistika – Kvalitativní znak

3D vs. 2D výsečový graf

Jaký je poměr velikosti výsečí A a C?

Jaký je poměr velikosti výsečí B a D?





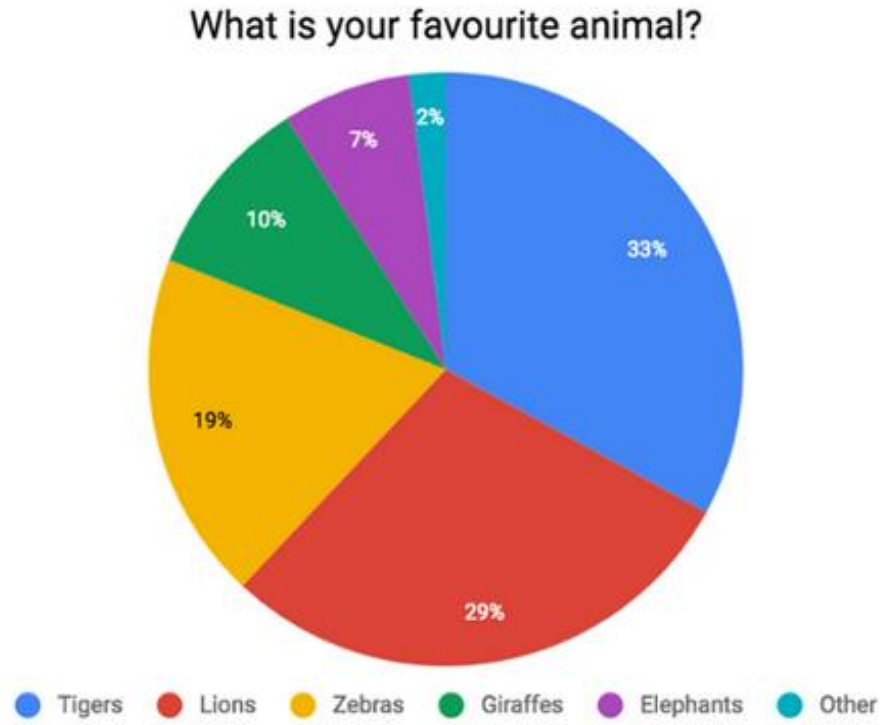
Popisná statistika – Kvalitativní znak

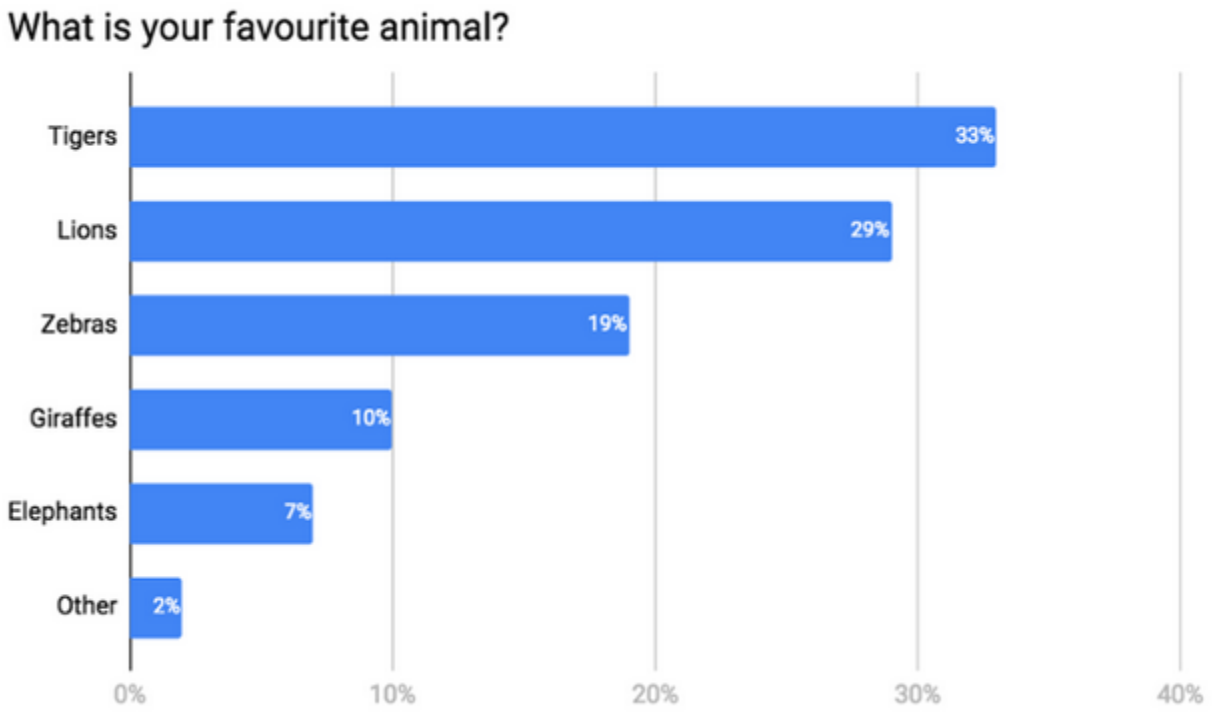
Jste pro navýšení hodinové dotace matematiky?

TAKHLE NE!!!

Nezapomínejte, že relativní četnosti byste měli uvádět pouze jako doplněk četností absolutních!

Je výsečový graf tou správnou volbou?



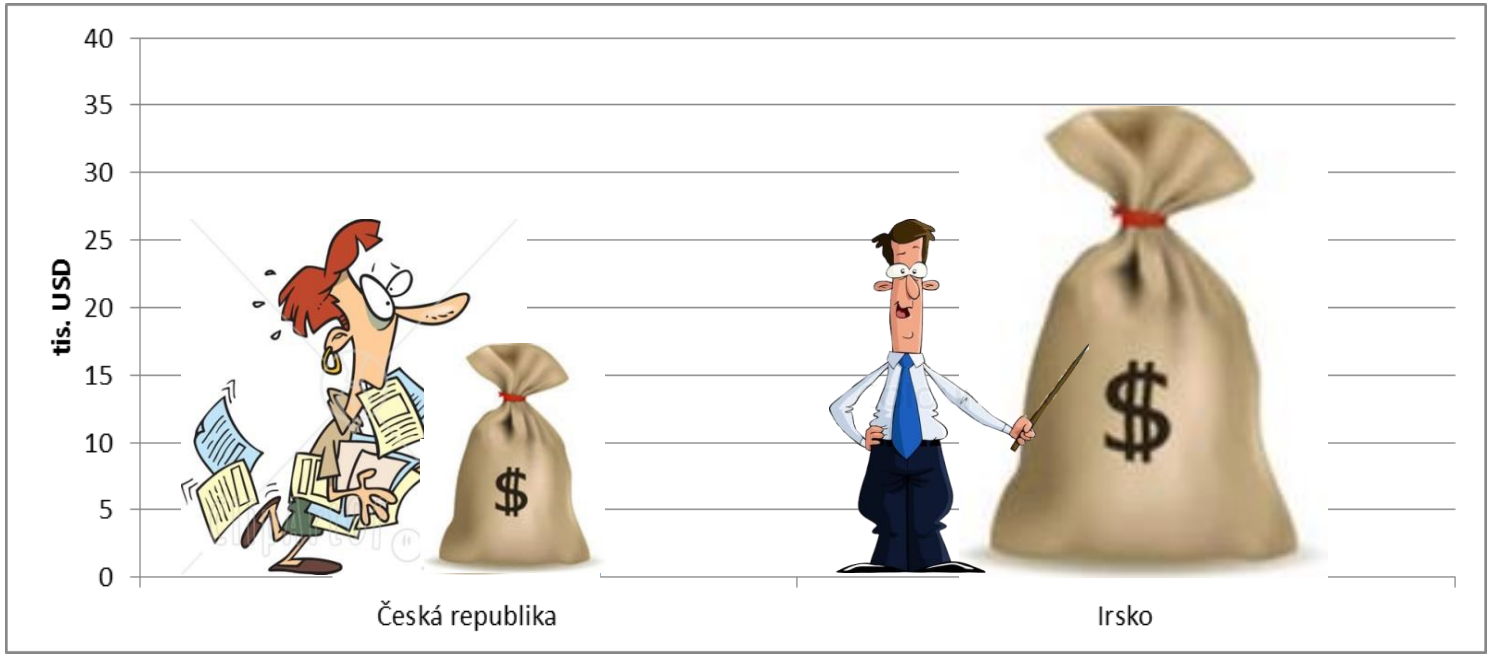


Jsou tygři populárnější než lvi?

Kolikrát jsou zebry populárnější než žirafy?

zdroj: https://blog.funnel.io/why-we-dont-use-pie-charts-and-some-tips-on-better-data-visualizations

Obrazkové grafy – užiteční pomocníci?

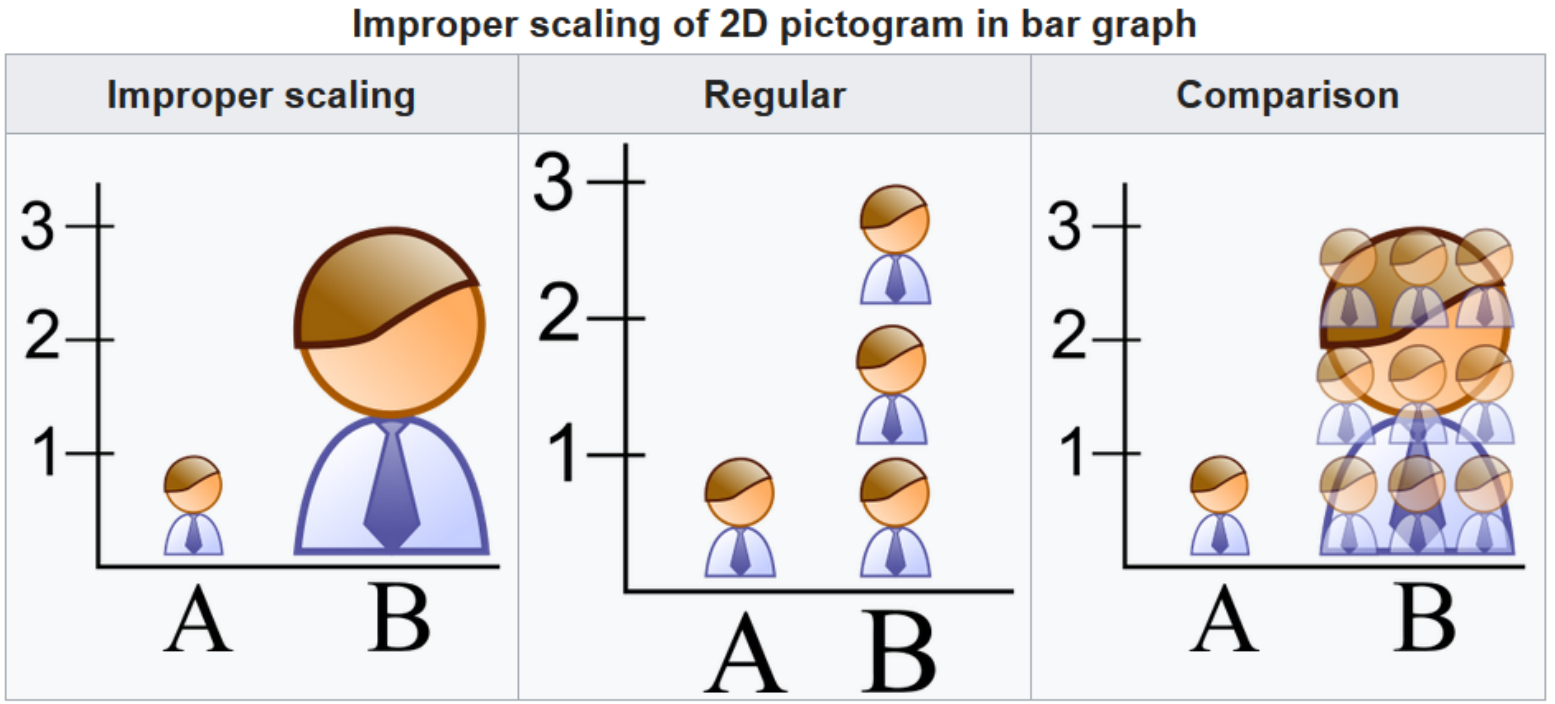


Srovnání průměrných ročních nástupních platů učitelů středních škol v ČR (17 244 $) a Irsku (34 604 $)

Obrazkové grafy – užiteční pomocníci?

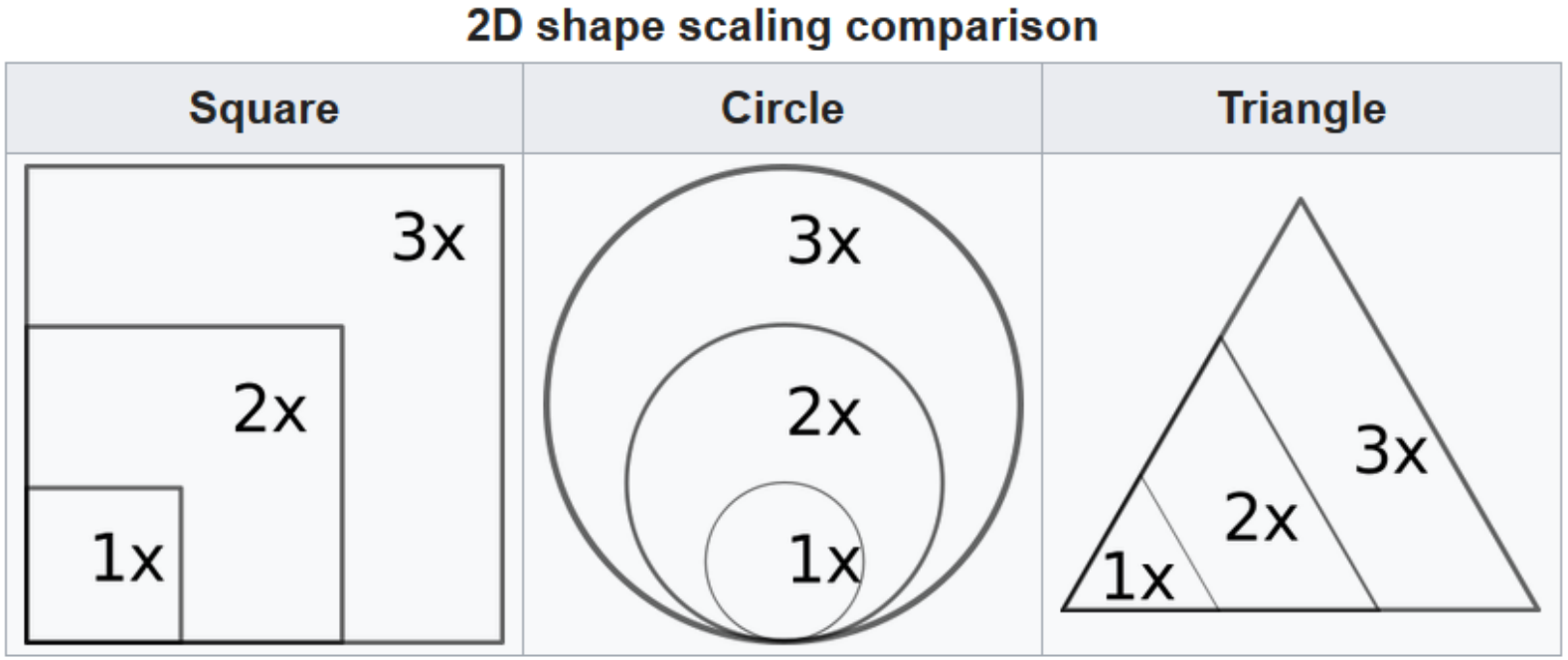
Srovnání průměrných ročních nástupních platů učitelů středních škol v ČR (17 244 $) a Irsku (34 604 $)

Obrazkové grafy – užiteční pomocníci?



Zdroj: UTTS, Jessica M. Seeing through statistics. 3rd ed. Belmont, CA: Thomson, Brooks/Cole, c2005. ISBN 0-534-39402-7

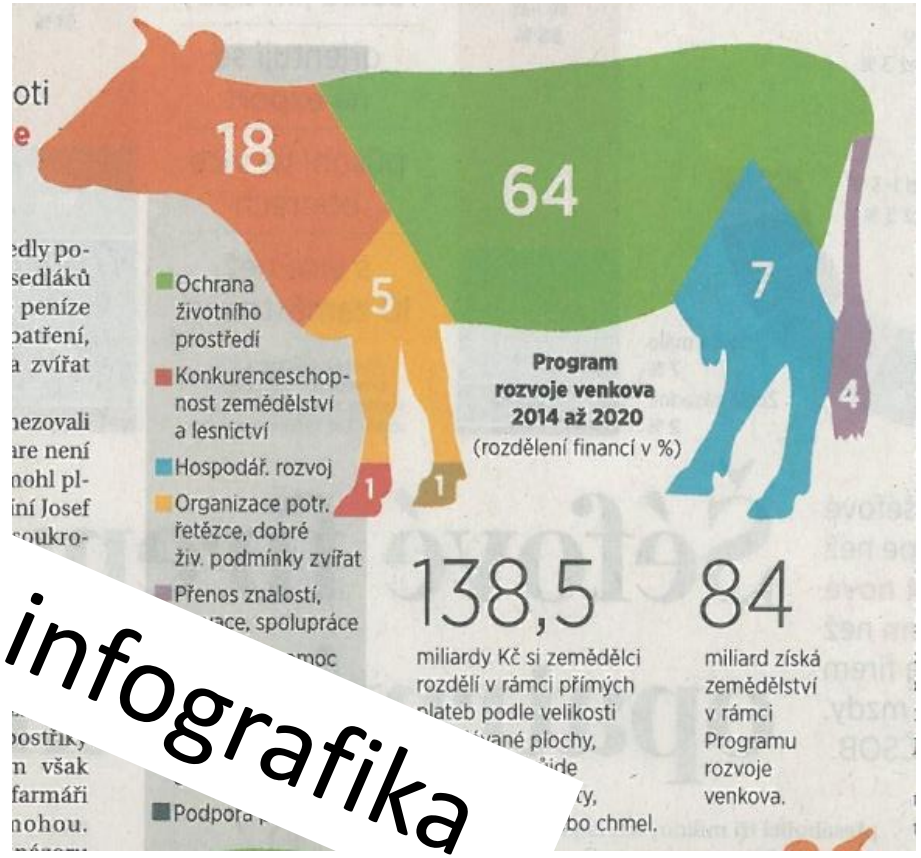
Obrazkové grafy – užiteční pomocníci?



Zdroj: UTTS, Jessica M. Seeing through statistics. 3rd ed. Belmont, CA: Thomson, Brooks/Cole, c2005. ISBN 0-534-39402-7

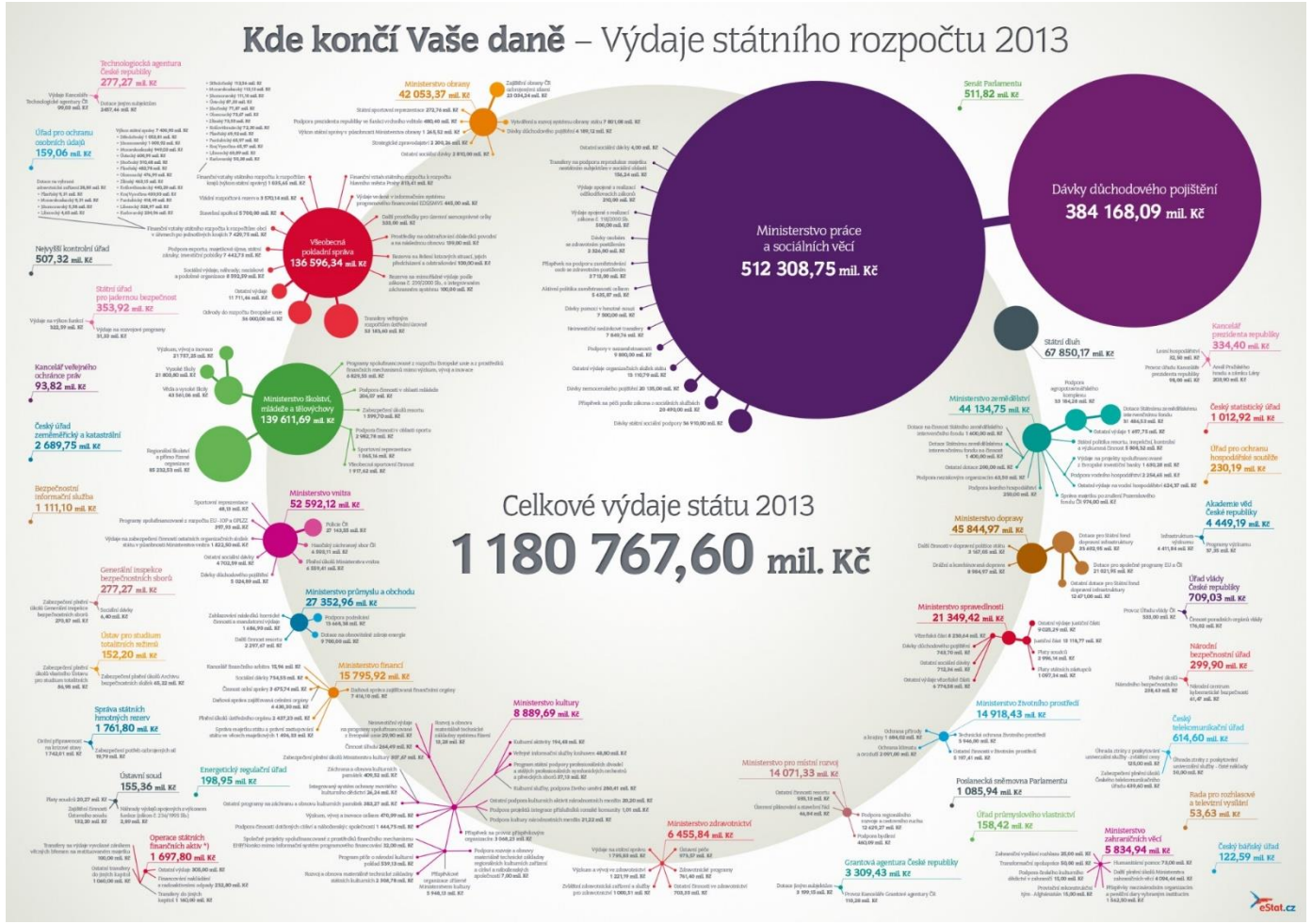
Pár příkladů z praxe

Obrazkové grafy – užiteční pomocníci?



Zdroj: Mf Dnes, 10. 7. 2014: Zemědělci si rozdělí miliardy. Krávy a vepři se budou mít lépe.

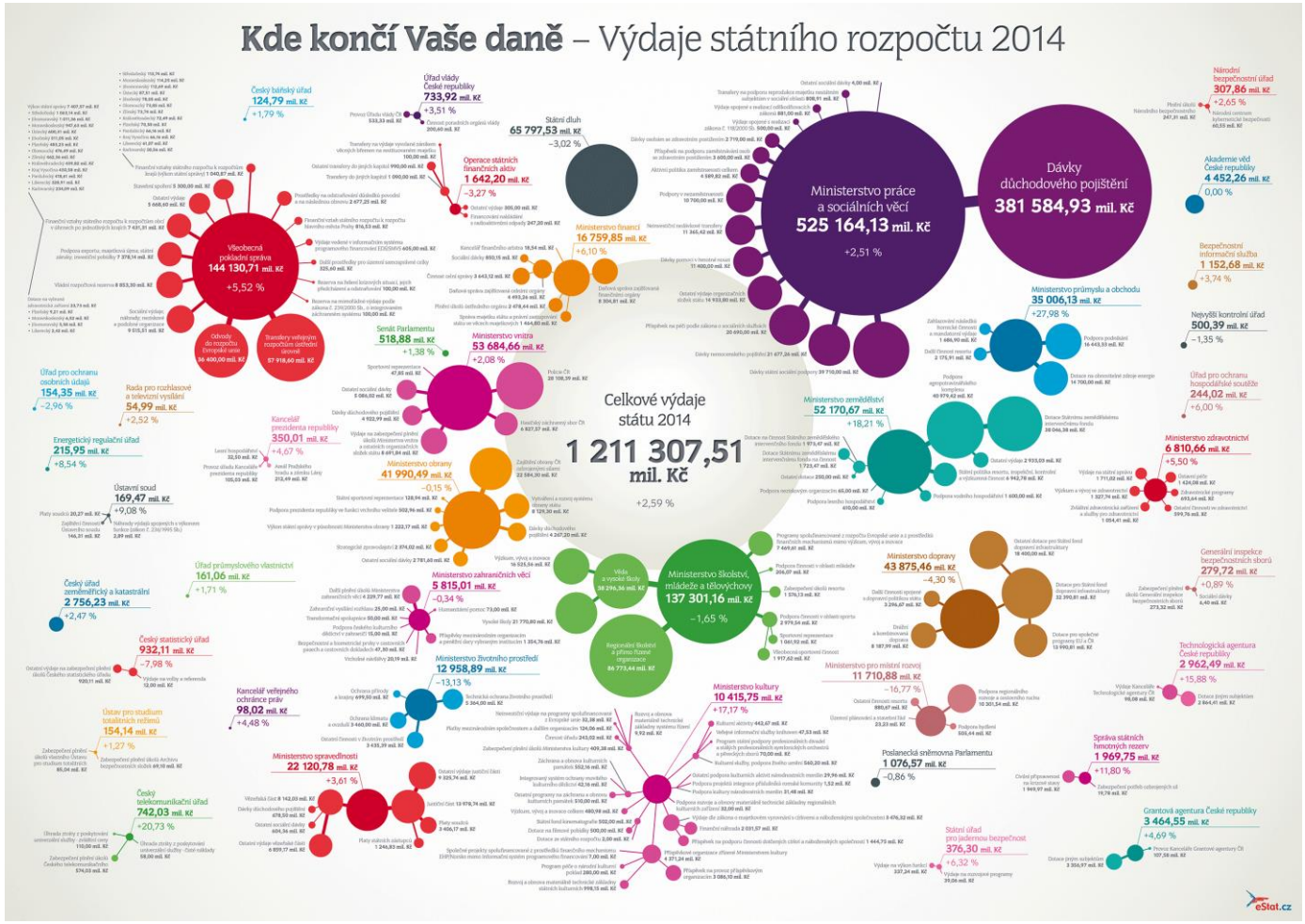
Obrazkové grafy – užiteční pomocníci?



„Úžasná infografika o výdajích státního rozpočtu České republiky v roce 2013“

Zdroj: http://www.estat.cz/zpravy/informace-k-projektum/kde-konci-vase-dane/

Obrazkové grafy – užiteční pomocníci?



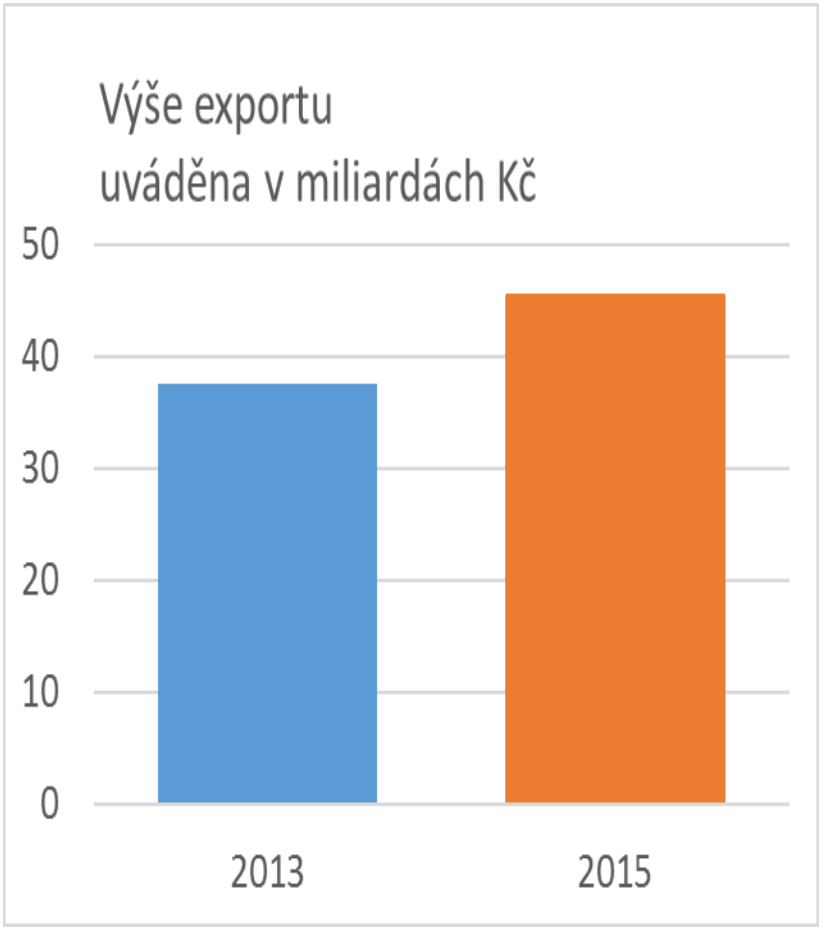
Zdroj: http://www.estat.cz/zpravy/informace-k-projektum/kde-konci-vase-dane/

Příklad s klobásou

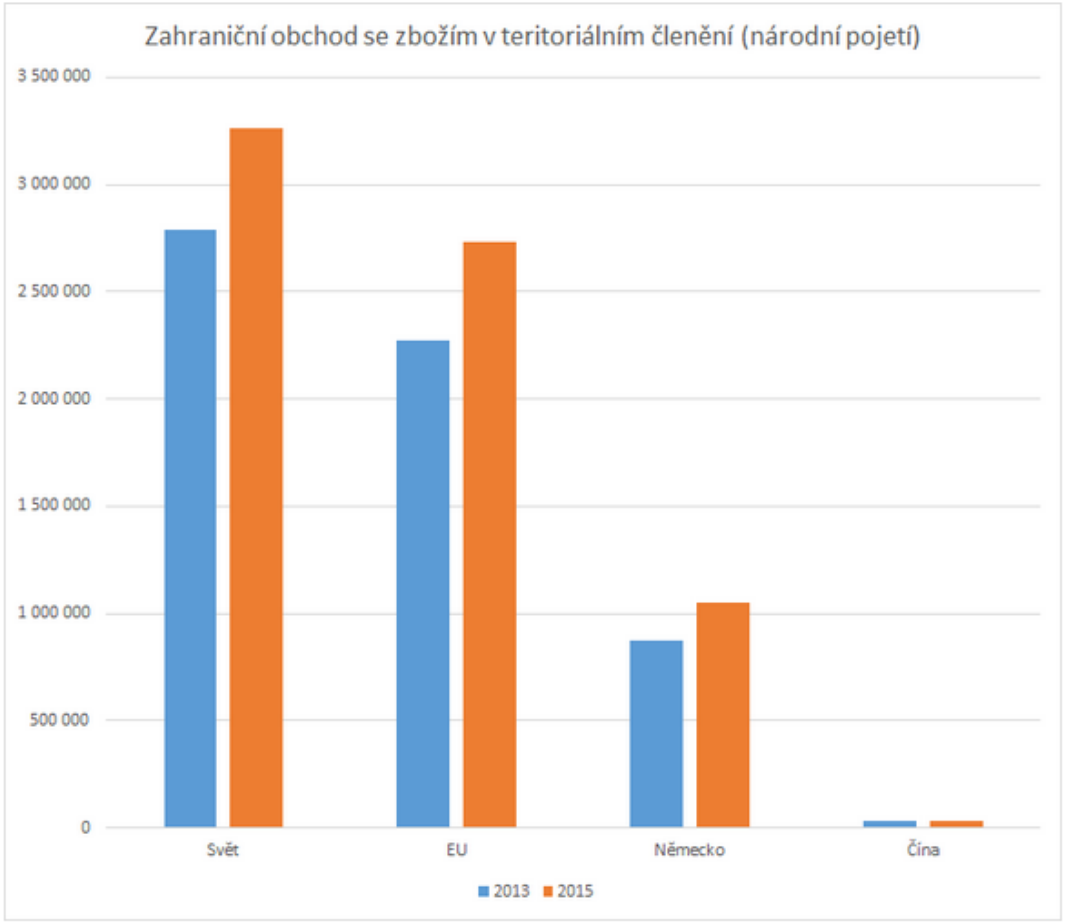
Příklad s klobásou

Český export do Číny aneb Porovnávejte!!!

Český export do Číny aneb Porovnávejte!!!



Český export do Číny aneb Porovnávejte!!!



Zdroj: https://www.souki.cz/kouzelne-grafy

Průzkum o představách studentů o budoucím zaměstnání

DNES MIMOŘÁDNÁ PŘÍLOHA

RADY Životopis a přijímací pohovor
Jak zaujmout budoucího zaměstnavatele a vyhnout se obvyklým chybám Str. C2 a C5

INTERNET Hledání práce přes LinkedIn
Pět tipů, jak mít atraktivní on-line profil a dostávat nabídky od personalistů Str. C5

ČTVRTEK 27. 3. 2014 WWW.IDNES.CZ C

Velký přehled trainee programů, které firmy připravily pro absolventy, najdete na str. C6 až C7

Ideální místo? Mezinárodní firma, nástup aspoň 18 tisíc

Podle kterých kritérií si absolventi škol vybírají práci, jaký by si přáli nástupní plat a nakolik jsou ochotni ustoupit ze svých požadavků, to zjišťoval nedávný průzkum. Některá čísla překvapila.

Jaké představy mají vysokoškoláci o své budoucí práci? To mapoval on-line průzkum

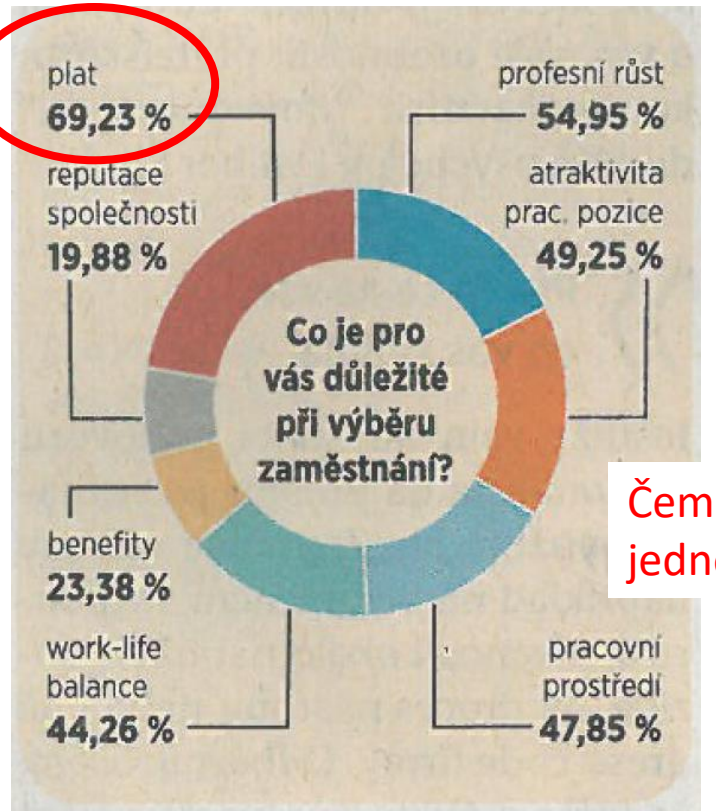
Zdroj: Společnost Studenta Media

***Zdroj**: Mimořádná příloha Mf Dnes, 27. 3. 2014 – výsledky šetření spol. Studenta Media (typ šetření: online dotazování, specifikace výběru: „přes tisíc vysokoškoláků ze všech ročníků po celé republice“)*

Průzkum o představách studentů o budoucím zaměstnání

S přesností na setinu procenta…

1 000 studentů … 100 %
10 studentů … 1 %
0,1 studentů … 0,01 %



Proč není součet 100 %?

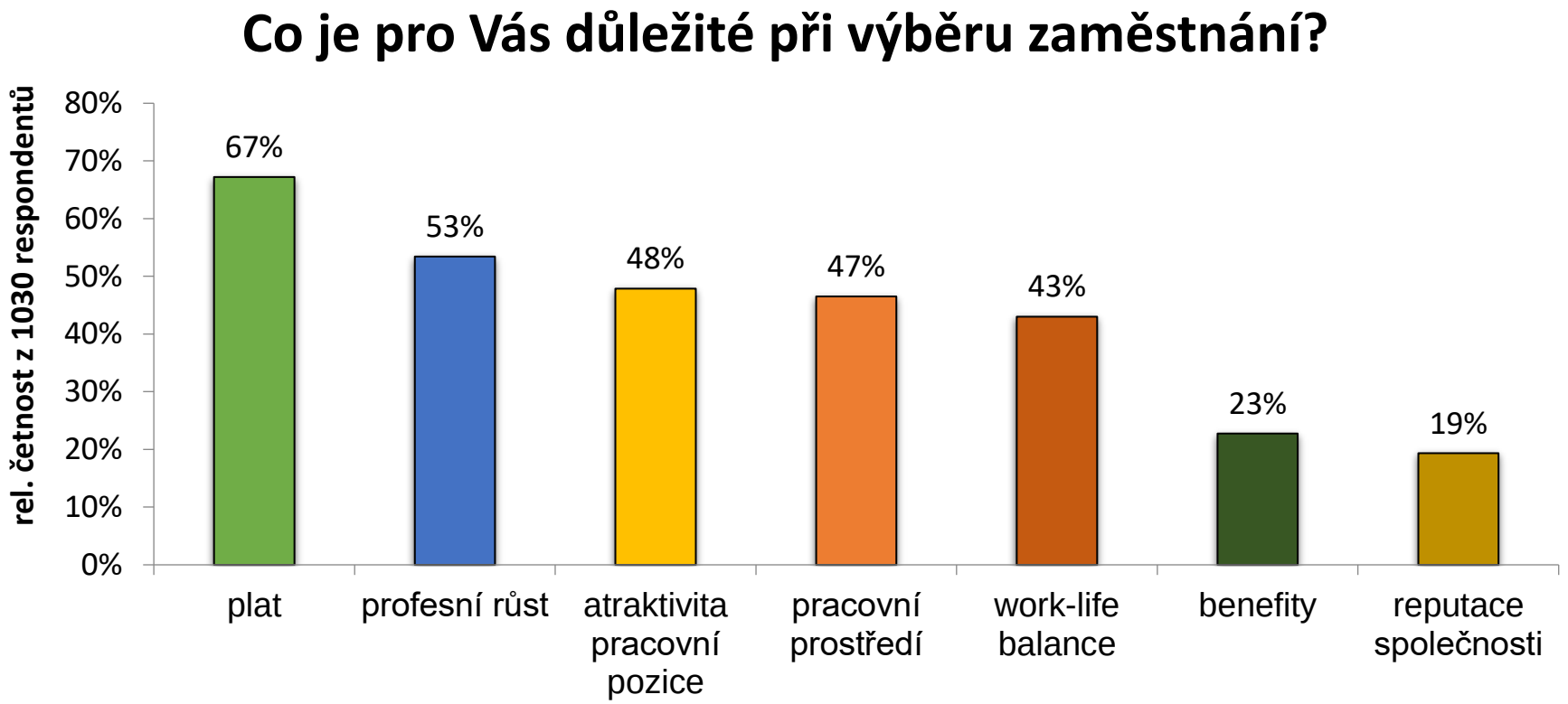
Čemu odpovídá velikost jednotlivých částí prstence?

***Zdroj**: Mimořádná příloha Mf Dnes, 27. 3. 2014 – výsledky šetření spol. Studenta Media (typ šetření: online dotazování, specifikace výběru: „přes tisíc vysokoškoláků ze všech ročníků po celé republice“)*

Jak výsledky šetření zobrazit správně?

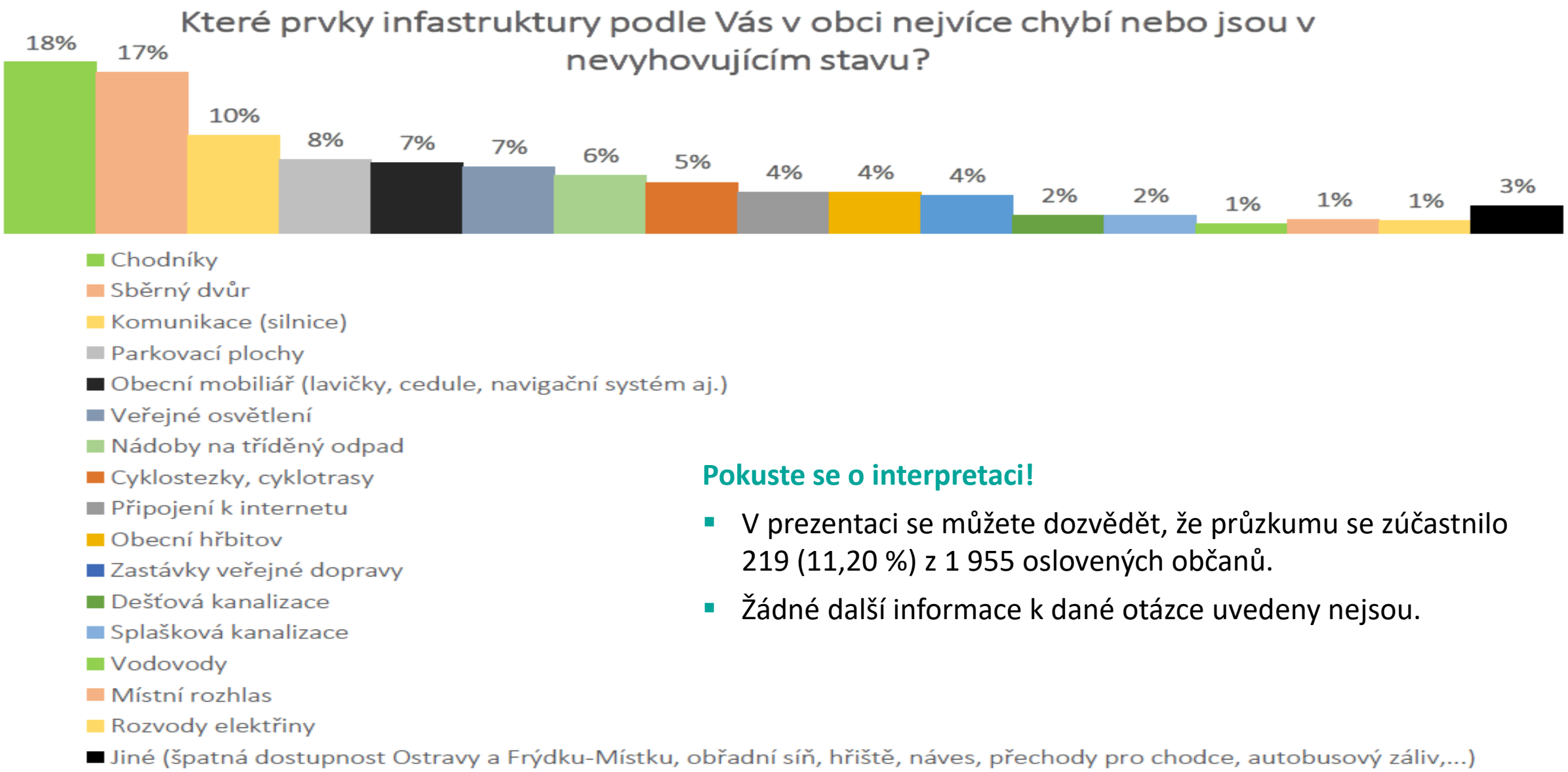
Co je pro Vás důležité při výběru zaměstnání? (vyberte 3 pro Vás nejdůležitější faktory)			
	četnost	rel. četnost (%)	rel. četnost (%) vzhledem k počtu respondentů
plat	692	22	67
profesní růst	550	18	53
atraktivita pracovní pozice	493	16	48
pracovní prostředí	479	16	47
work-life balance	443	14	43
benefity	234	8	23
reputace společnosti	199	6	19
celkem	**3 090**	**100**	**---**

Jak výsledky šetření zobrazit správně?



Barevné odlišení výplní sloupců není nutné (odborná publikace vs. marketing)!

Hodnocení dostupnosti infrastruktury obce



Pokuste se o interpretaci!

- V prezentaci se můžete dozvědět, že průzkumu se zúčastnilo 219 (11,20 %) z 1 955 oslovených občanů.
- Žádné další informace k dané otázce uvedeny nejsou.

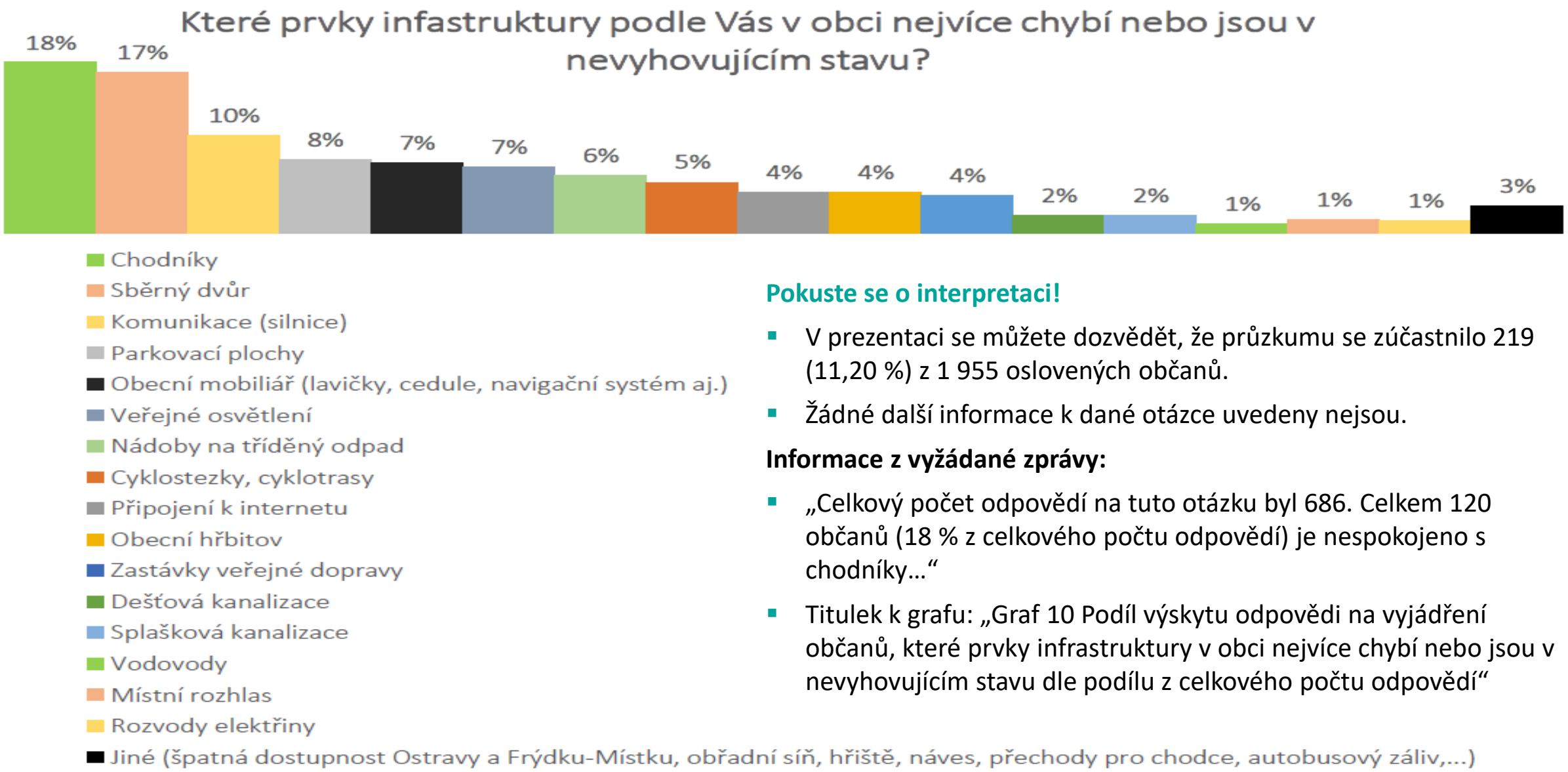
Zdroj: https://www.krmelin.cz/evt_file.php?file=1673&original=Dotaznikove_setreni_obce_Krmelin_vysledky_10_18.pdf (18. 11. 2018)

Informace z vyžádané podrobnější zprávy (na webu není zveřejněna)

Hodnocení infrastruktury obce a dopravy

Nevyhovující či chybějící prvky infrastruktury v obci dle občanů jsou uvedeny v grafu 10. Celkový počet odpovědí na tuto otázky byl 686. Celkem 120 občanů (18 % z celkového počtu odpovědí) je nespokojeno s chodníky v obci, dále 113 občanům (17 % z celkového počtu odpovědí) chybí sběrný dvůr a 71 občanů (10 % z celkového počtu odpovědí) je nespokojeno s komunikacemi (silnicemi) v obci, dále 53 občanů (8 % z celkového počtu odpovědí) s parkovacími plochami, 51 občanů (7 % z celkového počtu odpovědí) s obecním mobiliářem jako jsou lavičky, cedule, navigační systém aj. a 48 občanů (7 % z celkového počtu odpovědí) je nespokojeno s veřejným osvětlením. Všechny odpovědi jsou uvedeny v grafu 10. Občané

Hodnocení dostupnosti infrastruktury obce



Pokuste se o interpretaci!

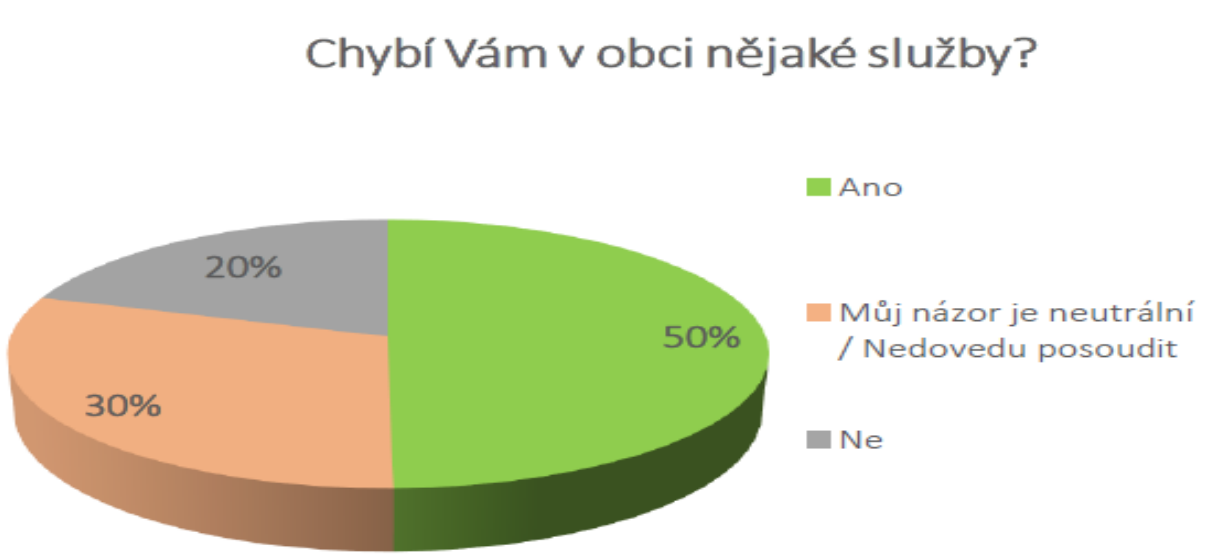
- V prezentaci se můžete dozvědět, že průzkumu se zúčastnilo 219 (11,20 %) z 1 955 oslovených občanů.
- Žádné další informace k dané otázce uvedeny nejsou.

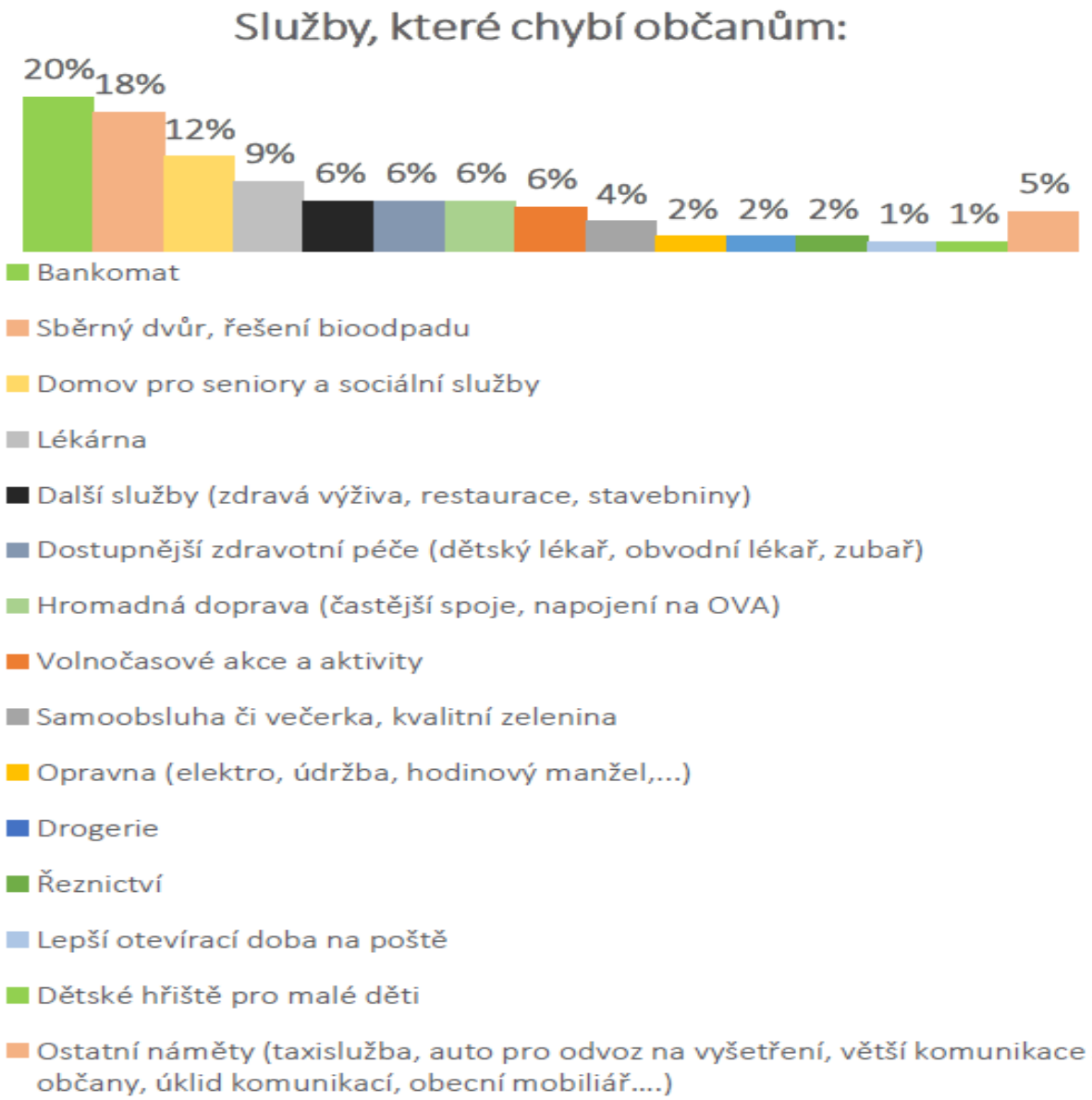
Informace z vyžádané zprávy:

- „Celkový počet odpovědí na tuto otázku byl 686. Celkem 120 občanů (18 % z celkového počtu odpovědí) je nespokojeno s chodníky…“
- Titulek k grafu: „Graf 10 Podíl výskytu odpovědi na vyjádření občanů, které prvky infrastruktury v obci nejvíce chybí nebo jsou v nevyhovujícím stavu dle podílu z celkového počtu odpovědí“

Zdroj: https://www.krmelin.cz/evt_file.php?file=1673&original=Dotaznikove_setreni_obce_Krmelin_vysledky_10_18.pdf (18. 11. 2018)

Hodnocení dostupnosti služeb





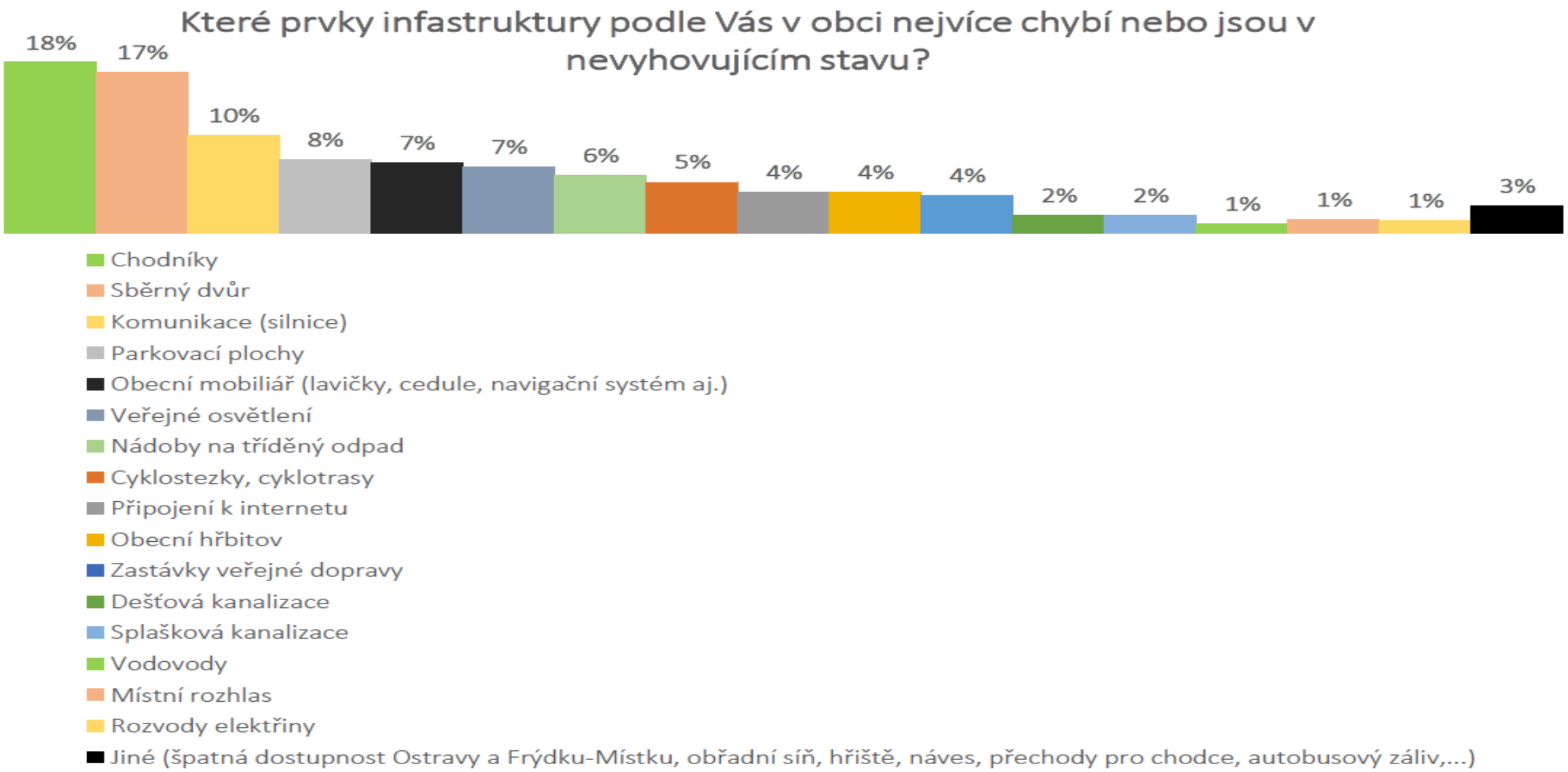
Co je pro respondenty „palčivějším“ problémem – stav chodníků nebo chybějící bankomat?

- V prezentaci se můžete dozvědět, že průzkumu se zúčastnilo 219 (11,20 %) z 1 955 oslovených občanů.
- Žádné další informace k dané otázce uvedeny nejsou.

Informace z vyžádané zprávy:

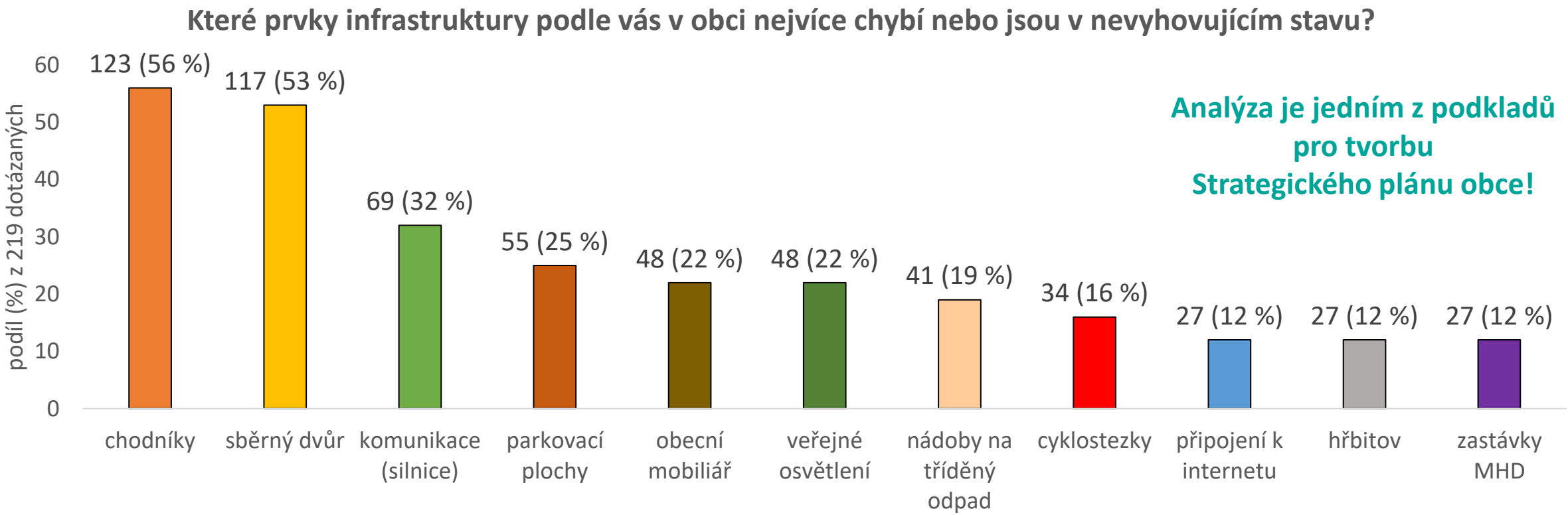
- „Občané uvedli celkem 160 námětů na chybějící služby v obci. Občanům nejčastěji chybí v obci bankomat (20 % z celkového počtu odpovědí, 32 odpovědí) …“
- Titulek k grafu: „Graf 6 Vyjádření občanů k otázce, jaké služby jim v obci chybí dle podílů z celkového počtu odpovědí “

Hodnocení dostupnosti infrastruktury obce



Zdroj: https://www.krmelin.cz/evt_file.php?file=1673&original=Dotaznikove_setreni_obce_Krmelin_vysledky_10_18.pdf (18. 11. 2018)

A co takto?



Analýza je jedním z podkladů pro tvorbu Strategického plánu obce!

Prvky infrastruktury, které označilo méně než 5 respondentů: dešťová a splašková kanalizace, vodovody, místní rozhlas, rozvody elektřiny, špatná dostupnost Ostravy a Frýdku-Místku, obřadní síň, hřiště, náves, přechody pro chodce, autobusový záliv

Děkuji za pozornost!

martina.litschmannova@vsb.cz

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY