

# Vizualizace pomocí hvězdiček a úskalí nelineárních vztahů

ONDŘEJ VENCÁLEK

Přírodovědecká fakulta UP, Olomouc

## Pár slov o vizualizaci dat a infografice

Přehledný obrázek je často lepší než mnoho slov, anebo než mnoho čísel. Informaci obsaženou v číslech (v datech) běžně vizualizujeme prostřednictvím nejrůznějších grafů. Vhodně zvolený graf pomůže příjemci informace (čtenáři, posluchači či divákovi) snadněji a rychleji pochopit informaci obsaženou v datech. Vizualizovat data tak, aby výsledný graf poskytl přehledné a snadno pochopitelné shrnutí dat, může být někdy velkou výzvou. A občas se může stát, že graf vyvolává zkreslenou či dokonce mylnou představu o realitě, kterou se snaží popsat. O tom, že tuto zkreslenou představu lze různými postupy vytvářet dokonce záměrně, psal v dnes již klasické populárně naučné knize *Jak lhát se statistikou* [1] už v polovině minulého století americký novinář a spisovatel Darrell Huff<sup>1)</sup>

Přestože nejdůležitějšími vlastnostmi informace jsou její pravdivost a relevance, je v době přemíry informací – a tou doba internetu a sociálních sítí zajisté je – klíčový způsob prezentace této informace a jeho schopnost zaujmout pozornost. Tradiční grafy, které samy o sobě mají za cíl usnadnit pochopení informace obsažené v datech, se stávají součástí moderních infografik kombinujících zajímavé grafické prvky (včetně obrázků, map apod.), stručné číselné informace, grafy a případně krátký text. Důraz je kladen na snadnou a rychlou pochopitelnost a zapamatovatelnost předkládané informace pro co nejvíce lidí.

Jsem přesvědčen, že tvorba různých infografik může být zajímavou činností v hodinách či kroužcích matematiky, informatiky, ale třeba i výtvarné výchovy, a to jak pro základní, tak i pro střední školy. Infografiky však mohou tematicky zasahovat např. do zeměpisu, přírodopisu, dějepisu, ale třeba také občanské či hudební výchovy. Inspirativní infografiky lze nalézt například v publikaci Českého statistického úřadu Česko v číslech [2]. Pře-

---

<sup>1)</sup>Na stránkách Českého statistického úřadu je k dispozici komentovaná česká vydání této knihy: <https://www.statistikaamy.cz/jak-lhat-se-statistikou/>.

svědčte se o jejím širokém využití ve výuce. A pokud vás a vaše žáky (studenty) tvorba infografik zaujme, můžete se přihlásit do *soutěže o nejlepší statistický plakát*<sup>2)</sup>, kterou každoročně organizuje právě Český statistický úřad. Přehledku dalších českých infografik z oblasti businessu, cestování, financí, gastronomie, internetu, telefonů, politiky, sportu a zdraví naleznete na stránce <https://ceskeinfografiky.cz/>. Za zmínku stojí také skutečnost, že v posledních letech pronikl trend infografik i do světa vědy, například v podobě takzvaných grafických abstraktů<sup>3)</sup>.

V tomto článku se soustředíme na atraktivní způsob vizualizace pomocí hvězdiček<sup>4)</sup>, který se tradičně používá k hodnocení kvality (v angličtině se používá spojení *star rating*). Pětihvězdičkový hotel je luxusní záležitostí nabízející vysokou kvalitu. Zisk tzv. Michelinské hvězdy<sup>5)</sup> (jedné až tří) je považován za vynikající hodnocení kvality restaurací, jímž se může pyšnit jen několik stovek restaurací po celém světě. Hvězdičkami jsme si však zvykli hodnotit „kvalitu“ v podstatě čehokoliv. Kolik hvězdiček z pěti dáte filmu, který jste nedávno zhlédli? Kolika hvězdičkami ohodnotíte ubytovací zařízení, ve kterém jste strávili dovolenou? A kolik hvězdiček z pěti dostane dnešní oběd ve školní jídelně?

Do podoby hvězdiček lze však převést v podstatě libovolnou číselnou veličinu, jejíž hodnoty porovnáваме, zejména pokud platí, že čím větší pozorované hodnoty jsou, tím lépe. *V tomto článku ukážeme, jak „převod čísel na hvězdičky“ provést. Zároveň si však ukážeme, jak zrádný takovýto převod může být, a jak velmi zavádějící (či dokonce účelově manipulativní) může být vizualizace pomocí hvězdiček.* Úvahy o různých způsobech převodu čísel na hvězdičky nás v závěru článku přivedou až k přemýšlení o povaze vztahu mezi různými veličinami.

### **Příklad: ilustrace efektivity vakcín pomocí hvězdiček**

Inspirací k sepsání tohoto textu byla infografika k článku o efektivitě vakcín proti onemocnění COVID-19 (viz obr. 1) publikovaném na webu

---

<sup>2)</sup> Informace o soutěži na stránce <https://csu.gov.cz/statisticky-plakat/domov>, kde jsou k dispozici rovněž ukázky prací z minulých ročníků.

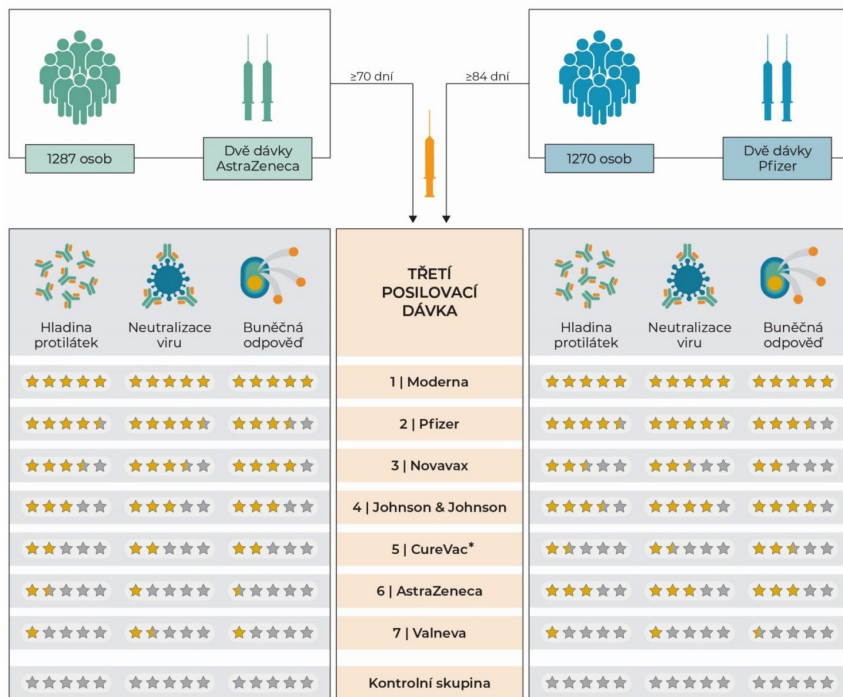
<sup>3)</sup> Viz např. stránka prestižního vydavatelství Elsevier zaměřeného na odborné publikace z mnoha různých oborů: <https://www.elsevier.com/authors/tools-and-resources/visual-abstract>.

<sup>4)</sup> Zajímavostí je, že nemusí jít vždy nutně o symboly „hvězdiček“, ale může jít o libovolné jiné symboly; např. litevská organizace atostogos kaime nabízející prázdninové ubytování na venkově používá místo hvězdiček piktogram čápa – symbolu venkovské krajiny této země, viz <https://www.countryside.lt/quality-ranking/>

<sup>5)</sup> [https://cs.wikipedia.org/wiki/Michelinův\\_průvodce](https://cs.wikipedia.org/wiki/Michelinův_průvodce).

Info-zdraví, což je web marketingové společnosti Mediaplanet zaměřený na poznatky a novinky z oblasti zdraví a nemocí, příběhy pacientů a nejnovější možnosti diagnostiky a léčby. Článek nazvaný „Studie zkoumala 7 boosterů proti covidu, rozdíl mezi nimi jsou značné“ [3] byl zveřejněn 9. září 2022 a českým čtenářům přinášel rekapitulaci britské studie publikované v časopise Lancet v prosinci 2021 [4]. Text českého článku byl doplněn infografikou, kterou připravila molekulární bioložka Lucia Ciglar<sup>6)</sup> věnující se popularizaci vědy mj. prostřednictvím tvorby infografik.

## POROVNÁNÍ POSILOVACÍCH VAKCÍN PROTI COVID-19



Zpracováno na základě studie: Munro, Alasdair PS et al. Safety and immunogenicity of seven COVID-19 vaccines as a third dose (booster) following two doses of ChAdOx1 nCov-19 or BNT162b2 in the UK (COV-BOOST): a blinded, multicentre, randomised, controlled, phase 2 trial. Lancet vol. 398 (2021). Data pro poloviční dávky vakcín Pfizer, Novavax a Valneva a údaje pro varianty delta a beta nejsou uvedeny.

science illustrated

Obr. 1 Infografika k článku o efektivitě boosterů proti covidu

<sup>6)</sup>Dr. Ciglar prezentuje výsledky své práce na stránkách <https://luciaciglar.com/> a <http://science-illustrated.com/>.

V infografice Lucie Ciglar byly číselné hodnoty tří různých veličin, které původní studie zkoumala (zkráceně jsou tyto veličiny nazvány *hladina protilátek*, *neutralizace viru* a *buněčná odpověď*), graficky vyjádřeny pomocí hvězdiček.

## Co lze vyčíst z infografiky o efektivitě vakcín

Věnujme se nejdříve interpretaci infografiky na obr. 1. Ve studii, jejíž výsledky tato infografika shrnuje, bylo porovnáváno celkem 16 skupin lidí. Všichni tito lidé byli vakcinováni dvěma nebo třemi dávkami vakcíny proti onemocnění COVID-19 – dvěma „základními“ a případně ještě jednou tzv. „posilovací“, tedy „boosterovou“. Polovinu z těchto skupin (8 skupin) tvořili lidé, kteří byli nejprve vakcinováni dvěma dávkami vakcíny společnosti AstraZeneca. Těch se týká levá část infografiky. Druhou polovinu skupin (rovněž 8 skupin) tvořili lidé, kteří byli nejprve vakcinováni dvěma dávkami vakcíny společnosti Pfizer. Těchto lidí se týká pravá část infografiky. Každá ze skupin je pak určena tím, jaká vakcína byla u ní použita coby booster. Bylo zkoumáno sedm různých boosterů (sedm skupin). Osmou skupinu, která je označena jako kontrolní, tvoří ti lidé, u nichž posilovací dávka nebyla aplikována, a byli tedy na rozdíl od ostatních očkovaní jen dvakrát.

U všech lidí ve studii byla změřena hladina protilátek (budeme se pro jednoduchost věnovat jen této ze tří sledovaných veličin). Přítomnost protilátek v lidském těle můžeme velmi zjednodušeně chápat jako jakousi připravenost daného jedince vyrovnat se s případnou nákazou.

Abychom se přesvědčili, že správně rozumíme obr. 1, zkusme zodpovědět následující čtyři otázky týkající se skupiny původně očkované vakcínou AstraZeneca (můžeme poté stejné otázky zodpovědět také pro skupinu očkovanou původně vakcínou od společnosti Pfizer):

1. Mají (ve skupině původně očkované vakcínou AstraZeneca) průměrně vyšší hladinu protilátek ti, kteří dostali booster od společnosti Pfizer, nebo ti, kteří dostali booster od společnosti AstraZeneca?
2. Který booster zajistil ve skupině původně očkované vakcínou AstraZeneca nejvyšší hladiny protilátek?
3. O kolik vyšší průměrnou hladinu protilátek má skupina s boosterem Moderna oproti skupině s boosterem Johnson and Johnson?
4. O kolik vyšší průměrnou hladinu protilátek má skupina s boosterem Valneva proti kontrolní skupině?

Tyto otázky se týkají jen srovnání v rámci skupin naočkovaných stejnou „základní“ vakcínou (AstraZeneca). Přidejme tedy nyní ještě tři otázky týkající se srovnání dvou velkých skupin určených typem vakcíny, kterou byli očkováni prvotně (AstraZeneca / Pfizer).

5. Platí, že lidé vakcinovaní nejprve vakcínou AstraZeneca a následně posilovací vakcínou Pfizer mají stejnou průměrnou hladinu protilátek jako lidé vakcinovaní nejprve vakcínou Pfizer a následně posilovací vakcínou Pfizer?
6. Ve které z následujících skupin byla zjištěna vyšší průměrná hladina protilátek? Skupina očkovaná nejprve vakcínou AstraZeneca a poté boosterem Johnson and Johnson, nebo skupina očkovaná nejprve vakcínou Pfizer a následně boosterem Novavax?
7. Ve které z následujících skupin byla zjištěna vyšší průměrná hladina protilátek? Skupina očkovaná nejprve vakcínou AstraZeneca a poté boosterem Valneva, nebo skupina očkovaná nejprve vakcínou Pfizer a následně již neočkovaná (kontrolní)?

Odpovědi na první dvě otázky jsou jednoduché. Ve skupině očkované nejprve vakcínou AstraZeneca má skupina boosterovaná vakcínou Pfizer 4,5 hvězdičky, zatímco skupina boosterovaná vakcínou AstraZeneca jen 1,5 hvězdičky. Vyšší hladině odpovídá více hvězdiček (správná odpověď je skupina boosterovaná vakcínou Pfizer). Nejvíce hvězdiček (pět) a tudíž i nejvyšší hladinu protilátek má ze všech skupin očkovaných vakcínou AstraZeneca skupina boosterovaná vakcínou Moderna.

Problémy však nastanou již u další dvojice otázek. Zatímco booster Moderna je ohodnocen 5 hvězdičkami, booster Johnson and Johnson jen 3 hvězdičkami. Ať pracujeme s poměrem 5 : 3 jakkoliv, nepovede se nám z něj zjistit, že ve skutečnosti je průměrná hladina ve skupině s boosterem Moderna více než pětinašobná oproti skupině s boosterem Johnson and Johnson. Skutečnou hodnotu lze zjistit z dat zveřejněných v originálním článku [4]. To, že poměr počtu hvězdiček nic nevyovídá o rozdílu mezi skupinami, je jasné ze čtvrté otázky, kde jedna ze srovnávaných skupin nemá ani jednu hvězdičku (poměr 1 : 0, jak víme, není matematicky definován). Navíc by zde mohlo dojít k mylné interpretaci, že hladina protilátek v kontrolní skupině je nulová. Ve skutečnosti je průměrná hladina ve skupině s boosterem Valneva přibližně dvojnásobná oproti kontrolní skupině, což ovšem z hvězdičkové infografiky vyčíst nelze.

Jestliže nám u otázek 3 a 4 postupně došlo, že na ně neumíme na základě srovnání počtu hvězdiček uspokojivě odpovědět, u otázek 5 až 7 reálně hrozí, že na základě srovnání počtu hvězdiček odpovíme nesprávně. Otázka 5: ačkoliv mají obě skupiny 4,5 hvězdičky, je ve skupině, která byla nejprve očkována vakcínou Pfizer, průměrná hladina protilátek téměř o 40 % vyšší. Správná odpověď na otázku 5 je tedy záporná. U další otázky má sice skupina očkováná nejprve vakcínou AstraZeneca a poté boosterem Johnson and Johnson 3 hvězdičky, zatímco skupina očkováná nejprve vakcínou Pfizer a následně boosterem Novavax jen 2,5 hvězdičky, přesto je průměrná hladina v druhé ze srovnávaných skupin téměř dvojnásobná oproti skupině první. Skupina očkováná nejprve vakcínou AstraZeneca a poté boosterem Valneva (jedna hvězdička) má ve skutečnosti průměrně méně protilátek než skupina očkováná nejprve vakcínou Pfizer a následně již neočkováná (nula hvězdiček). Infografika ve skutečnosti neumožňuje srovnání levé a pravé části, tedy skupin s různým počátečním očkováním. Proč tomu tak je? Hvězdičky jsou totiž „udíleny“ v těchto skupinách odděleně. Jednotlivé skupiny (boostery) jsou vždy srovnávány s kontrolní skupinou, ovšem se stejným počátečním očkováním. A protože kontrolní skupina očkováná vakcínou Pfizer má výrazně vyšší hladiny protilátek než kontrolní skupina očkováná vakcínou AstraZeneca, dochází k výše uvedeným „paradoxům“.

Názorně jsme tedy předvedli, k jakým netušeným problémům může při interpretaci zdánlivě jednoduché infografiky docházet.

## **Jak převést čísla na hvězdičky – matematické minimum**

Jak vůbec převést čísla na „hvězdičky“? V předchozím textu jsme se při hledání odpovědi na otázky 3 a 4 zamýšleli nad opačným problémem – potřebovali jsme se na základě počtu hvězdiček dozvědět něco o původních číselných hodnotách. Zatímco porovnání počtu hvězdiček (otázky 1 a 2) nám umožnilo srovnat velikost číselných hodnot (hladin protilátek), detailnější informace (odpovědi na otázky typu „o kolik“ či „kolikrát“ více) zůstávají bez znalosti toho, jak byla čísla převedena na hvězdičky, nedostupné.

Mějme na začátku číselné hodnoty, které chceme vizualizovat. V našem případě půjde o (průměrné) hladiny protilátek ve srovnávaných skupinách. Pro jednoduchost budeme pracovat jen se skupinami, které byly prvotně očkovány vakcínou AstraZeneca. Hladiny protilátek změřené v jednotlivých skupinách jsou v původním článku vztaheny k průměrné hladině protilátek

v referenční (kontrolní) skupině. Pro jednotlivé skupiny jsou tyto hodnoty rovny 32,30; 24,48; 8,75; 5,84; 5,05; 3,25; 2,20; 1. Znamená to, že v první skupině (booster Moderna) byla průměrná hladina protilátek 32krát vyšší oproti kontrolní skupině, ve druhé skupině přibližně 24krát větší atd. Poslední hodnota je rovna 1, neboť se týká samotné kontrolní skupiny.

Nyní je třeba rozmyslet si, jakou škálu hvězdiček použijeme. Zvolíme, kolika hvězdičkám bude odpovídat nejvyšší hodnota. Zůstaneme u volby 5 hvězdiček, jako v infografice na obr. 1. Dále je třeba rozmyslet, zda nejmenší možný počet hvězdiček je 0, nebo 1. Zde už můžeme diskutovat výhody a nevýhody obou možností. Bylo by dobré, aby „0 hvězdiček“ odpovídalo nulové hladině protilátek. Obhajitelná je však i možnost, kde „0 hvězdiček“ odpovídá „nulovému efektu boosteru“, tedy kontrolní skupině. Pak by však měla být tato skutečnost jasně deklarována, aby nedocházelo k nedorozumění. Poslední volbou je rozhodnutí, zda budeme uvažovat jen „celé hvězdičky“, nebo zda dovolíme i „půlhvězdičky“, jako je tomu na obr. 1. Shrňme tedy nyní možné volby pro vyjádření počtu hvězdiček:

$$H_0 = \{0, 1, 2, 3, 4, 5\}$$

$$H_1 = \{1, 2, 3, 4, 5\}$$

$$H_2 = \{0, 0,5, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5\}$$

$$H_3 = \{1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5\}$$

Počet hvězdiček příslušný jednotlivým číselným hodnotám zjistíme pomocí nějaké funkce, která číselným hodnotám z množiny

$$D = \{1, 2,20, 3,25, 5,05, 5,84, 8,75, 24,48, 32,30\}$$

přiřazuje hodnoty z množiny  $H_0$  (případně  $H_1$ ,  $H_2$  či  $H_3$ ). Budeme tuto funkci označovat písmenem  $f$ . Intuice nám velí, aby šlo o funkci rostoucí, tedy takovou, která vyšším číselným hodnotám z množiny  $D$  přiřazuje vyšší hodnoty z množiny  $H_0$ . Jelikož obor hodnot (množina  $H_0$ ) má méně prvků než množina  $D$ , bude požadavek tzv. ryzí monotonie nespílitelný, a proto si vystačíme s požadavkem, aby funkce byla neklesající. Uvědomme si však, že toto je v podstatě jediné omezení, které na hledanou funkci klademe. V následující části uvedeme několik možných funkcí použitelných pro zobrazení množiny  $D$  do množiny  $H_0$ .

Připomeňme, že výše uvedené úvahy se dají matematicky vyjádřit pomocí rovností

$$f(32,3) = 5, \tag{1}$$

$$f(1) = 1, \quad \text{případně } f(1) = 0. \tag{2}$$

## Příklady transformačních funkcí

### *Transformace pomocí lineární funkce*

V tomto případě uvažujeme funkci  $f^*$  ve tvaru

$$f^*(x) = ax + b,$$

kde  $a$  a  $b$  jsou parametry, jejichž hodnotu můžeme vypočítat dosazením z rovností (1) a (2). Tak dostaneme soustavu dvou rovnic o dvou neznámých  $a$  a  $b$ .

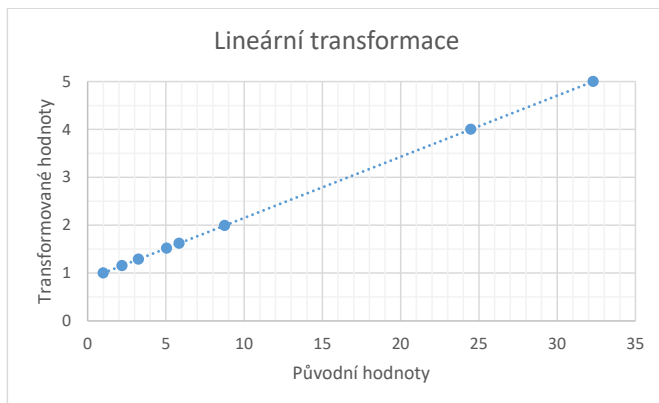
$$5 = a \cdot 32,3 + b \tag{3}$$

$$1 = a \cdot 1 + b$$

Řešením této rovnice dospějeme k následující podobě lineární funkce:

$$f^*(x) = 0,128x + 0,872 \tag{4}$$

V grafu této funkce znázorněném na obr. 2 jsou vyznačeny body, jejichž  $x$ -ová souřadnice odpovídá hodnotám z množiny  $D$ .



Obr. 2 Lineární transformace na interval  $(1; 5)$



Aby obor hodnot odpovídal množině  $H_0$ , je třeba ještě výslednou hodnotu zaokrouhlit na celé číslo. Poznamenejme, že při požadavku na širší obor hodnot (např.  $H_2$  či  $H_3$ ) bychom zaokrouhlovali k nejbližšímu celočíselnému násobku hodnoty 0,5.

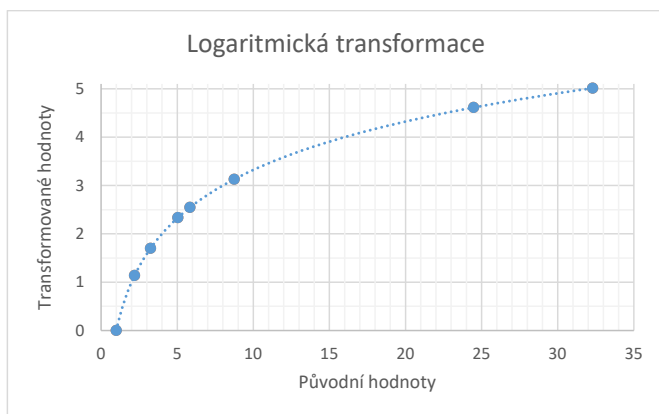
Hodnocení hladin protilátek vyjádřené pomocí hvězdiček při použití této (lineární) transformace je na obr. 6 vlevo nahoře. Jak se toto hodnocení změní, když místo podmínky  $f(1) = 1$  budeme uvažovat podmínku  $f(1) = 0$ , je vidět na obr. 6 vpravo nahoře. V obou případech je vidět velký „odstup“ dvou nejvyšších hodnot od ostatních.

### ***Transformace pomocí logaritmické funkce***

Velký odstup dvou největších hodnot od ostatních, patrný při použití lineární funkce, poněkud znesnadňuje porovnání šesti zbylých skupin – např. v situaci na obrázku 6 vpravo nahoře je těchto šest skupin „natěsnáno“ v rozmezí 0 až 1 hvězdička. Pokud bychom chtěli počty hvězdiček rozprostřít rovnoměrněji, mohli bychom využít logaritmickou transformaci. Pro snadnou interpretovatelnost zvolíme logaritmus o základu 2. Uvažujme tedy funkci

$$f^*(x) = \log_2(x).$$

Graf této funkce spolu se zvýrazněnými body, jejichž  $x$ -ové souřadnice odpovídají hodnotám z množiny  $D$ , je znázorněn na obr. 3. Z tvaru křivky je patrný efekt zmenšení mezery mezi dvěma nejvyššími a zbylými hodnotami.



Obr. 3 Logaritmická transformace

Význam hvězdičky nám objasní následující jednoduchá úvaha:

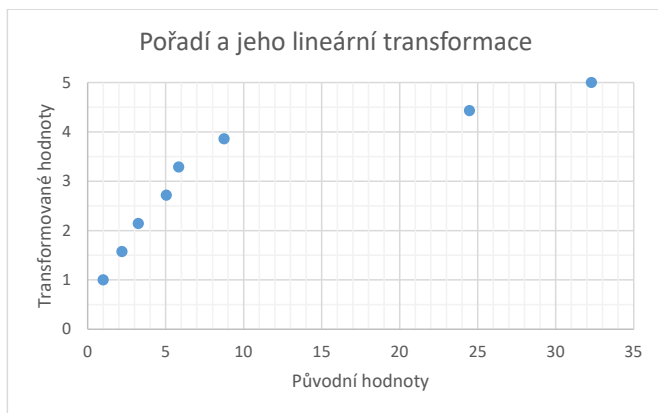
$$f^*(2x) = \log_2(2x) = \log_2(2) + \log_2(x) = 1 + \log_2(x) = f^*(x) + 1. \quad (5)$$

Z rovnice (5) plyne, že hodnota funkce  $f^*$  se zvětší o 1 (přibude 1 hvězdička), když se hodnota  $x$  zdvojnásobí. Každá hvězdička navíc tedy znamená zdvojnásobení hladiny protilátek. Pro úplnost dodejme, že je opět třeba výsledné hodnoty zaokrouhlovat.

Problémem použití této transformace je nutnost pečlivě vysvětlit interpretaci výsledného zobrazení.

### ***Transformace pomocí pořadí a jeho lineární transformace***

Snad úplně nejjednodušší možností je použití pořadí jednotlivých hodnot. Osmi hodnotám z množiny  $D$  přiřadíme jejich pořadí v uspořádání od nejmenší po největší hodnotu. Tedy číslo 1 přiřadíme nejmenší hodnotě, číslo 2 druhé nejmenší atd. Největší hodnotě (z osmi) pak bude přiřazeno číslo 8. Kdybychom měli k dispozici osm hvězdiček, mohli bychom je přidělovat přímo podle pořadí. Pokud se však chceme pohybovat v rozmezí 1 až 5 hvězdiček, musíme provést ještě lineární transformaci intervalu  $\langle 1; 8 \rangle$  na interval  $\langle 1; 5 \rangle$ . Výslednou transformaci (před závěrečným zaokrouhlením hodnot) znázorňuje graf na obr. 4.



Obr. 4 Transformace pomocí pořadí a jeho lineární transformace

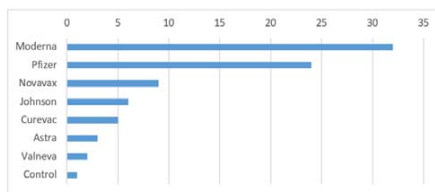
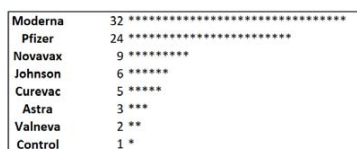
Výslednou reprezentaci dat z množiny  $D$  pomocí hvězdiček pak zobrazuje graf na obr. 6 vpravo dole. Uvědomme si, že použití této transformace

je rezignací na možnost odpovídat na otázky typu *o kolik* se jednotlivé skupiny liší. Zobrazeno je jen pořadí skupin bez rozlišení velikosti rozdílu mezi nimi.

## Úskalí vizualizace pomocí hvězdiček a možné řešení

Odhlédněme nyní od technických detailů jednotlivých transformací popsaných výše a soustředíme se na výsledné vizualizace znázorněné na obr. 6. Zopakujme, že jde o vizualizace téže datové sady (čítající osm hodnot). Přitom si povšimněme, jak velmi různé tyto vizualizace jsou. Každá z těchto možností je obhajitelná. Přesto výsledný dojem, který z nich člověk získá o efektivitě jednotlivých boosterů, je dosti odlišný. Pro úplnost dodejme, že ani jedna z navržených možností není přesně shodná s vizualizací na obr. 1.

Problémy s vizualizací do jisté míry vznikly tím, že jsme dopředu omezili maximální počet hvězdiček na pět. Kdybychom se takto neomezili, mohli bychom každé skupině „přiznat“ tolik hvězdiček, jaká je v ní (po zaokrouhlení) průměrná hodnota sledované veličiny (hladiny protilátek). Výsledný graf by pak vypadal tak, jak je ukázáno na obr. 5 vlevo. Od něj už je pak jen krok k použití standardního grafického nástroje – sloupcového grafu, viz obr. 5 vpravo.

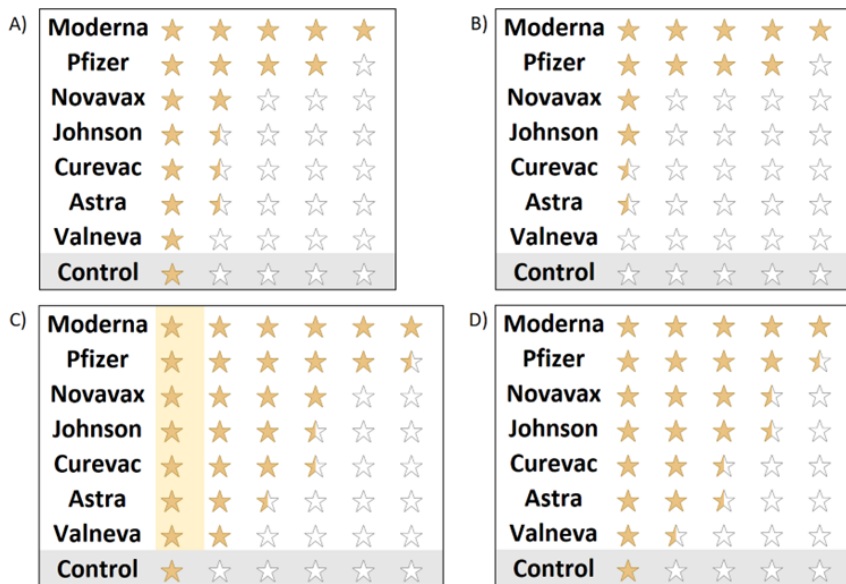


Obr. 5 Vizualizace pomocí vícero hvězdiček (vlevo) a pomocí histogramu (vpravo)

## Úvahy o povaze vztahu mezi dvěma veličinami

V článku jsme se zabývali tím, jak číselné hodnoty vyjadřující hladinu protilátek převést na počet hvězdiček, který obvykle chápeme jako vyjádření kvality čehosi, např. kvality ochrany před nákazou. Tichým předpokladem přitom je, že „čím více hvězdiček, tím lépe“. Jakou povahu má však vztah mezi studovanými veličinami, tedy v našem případě mezi množstvím protilátek a mírou ochrany proti nákaze? Jde o přímou úměru? Znamená opravdu dvojnásobné množství protilátek dvojnásobnou ochranu

proti nemoci? A stonásobné stonásobnou? Znamená nulové množství protilátek nulovou ochranu? Stačí se zeptat imunologů a zjistíme, že nulová hladina protilátek neznamená nulovou ochranu před infekcí. Vyjádříme-li míru ochrany před infekcí pomocí procent, pak tato nemůže být lineární funkcí množství protilátek, neboť každá přímka s kladnou směrnici po čase překročí hodnotu 100 %. Povaha závislosti mezi množstvím protilátek a mírou ochrany je tedy složitější. Věřím, že inspirování výše uvedenými příklady transformačních funkcí jsme připraveni uvažovat o mnoha různých možných podobách této závislosti. Jak zjistit, jaká je povaha tohoto vztahu ve skutečnosti? To je na další dlouhé pojednání.



Obr. 6 Vizualizace průměrných hodnot hladiny protilátek ve skupinách s různými typy boosteru dle studie [4]. Počet hvězdiček stanoven za použití lineární transformace s podmínkou  $f(1) = 1$  (vlevo nahoře), resp. s podmínkou  $f(1) = 0$  (vpravo nahoře), logaritmické transformace (vlevo dole) a pořadí (vpravo dole).

## Závěr

V tomto příspěvku jsme se věnovali atraktivní možnosti vizualizace číselných hodnot pomocí „hvězdiček“ (či jiných symbolů) běžně využívaných k hodnocení kvality na předem dané škále. Pomocí sedmi otázek, které jsme

čtenářům položili, jsme ukázali, že takovéto vizualizace mohou snadno vést k mylné interpretaci. Rovněž jsme ukázali, jak moc rozdílným vizualizacím jedněch dat můžeme dojít volbou transformační funkce, která číselným hodnotám přiřazuje počet hvězdiček. Citlivostí na tuto volbu se celá procedura stává náchylnou k nejrůznějším manipulacím. Dodejme, že náchylnost k manipulacím se netýká jen grafů využívajících hvězdiček. Jde o daleko širší problém, jehož existenci je dobré mít na paměti.

## Dodatek

Můžete samostatně zkusit provést stejnou vizualizaci pomocí hvězdiček, jaká je na obr. 6. Tento technický dodatek nabízí poměrně detailní popis postupu, který jsme užili my. Vizualizace byly provedeny pomocí softwaru excel. Čtenáři si mohou stáhnout soubor (<https://mfi.upol.cz/files/34/3402/Vencalek-vizualizace.xlsx>), v němž jsou provedeny všechny transformace popsané v tomto článku. Detailně budeme komentovat lineární transformaci, která by čtenářům mohla být nejbližší:

- Číselné hodnoty, ze kterých vycházíme a které jsou uvedeny výše v tomto článku, zapíšeme pod sebe do jednoho sloupce (sloupec C). Největší hodnota (32,3) nechť je v prvním řádku (řádek 15), nejnižší hodnota (1,0) nechť je poslední (řádek 22).
- Nyní definujeme sloupec D, ve kterém budou hodnoty funkce  $f^*(x)$ :
  - Nejprve zadáme do buňky B15 hodnotu, na kterou se transformuje nejvyšší hodnota  $x$ , tj. 5, a do buňky D22 hodnotu, na kterou se transformuje nejnižší hodnota, tj. 1.
  - Následně vypočteme hodnotu směrnice  $a$  a absolutního členu  $b$ , vyplývající ze soustavy rovnic (3). Zapišme například hodnotu parametru  $a$  do buňky G15 a hodnotu parametru  $b$  do buňky H15. Ze soustavy rovnic (3) plyne, že hodnotu  $a$  získáme jako podíl  $(D15-D22)/(C15-C22)$  a hodnotu  $b$  jako  $D22-G15*C22$ .
  - Nyní se vrátíme k zatím nevyplněným buňkám ve sloupci D. V nich potřebujeme dosadit hodnoty ze sloupce C za  $x$  v rovnici (4). Můžeme tedy začít v buňce D16, kam vepíšeme vzorec  $=G15*C16+H15$ . Ten pak zkopírujeme do buněk D17 až D21.
- V dalším kroku je třeba hodnoty ze sloupce D zaokrouhlit na celá čísla, či celočíselné násobky hodnoty 0,5, a tím získat počet hvězdiček. K tomu můžeme použít funkci ZAOKROUHLIT. My jsme

použili vzorec  $=\text{ZAOKROUHLIT}(D22*2;0)/2$  Tím jsme získali zao-  
krouhlení umožňující „půlhvězdičky“.

- Hvězdičky budeme vykreslovat v dalších pěti sloupcích, v našem pří-  
padě to budou sloupce T až X. Do prvního z nich (sloupec T) pouze  
zkopírujeme hodnoty ze sloupce E, tedy počty hvězdiček. Další slou-  
pec (U) bude obsahovat hodnoty ze sloupce T zmenšené o 1. Mohli  
bychom ho tedy chápat jako počet hvězdiček, které zbývají vykreslit.  
V dalším sloupci opět hodnoty ze sloupce U zmenšíme o 1 atd. Ně-  
které z hodnot mohou být i záporné. Takto vyplníme všechny buňky  
v rozmezí T15 až X22.
- Označíme oblast T15 až X22 a v nabídce *Domů – Styly* zvolíme *Pod-  
míněné formátování* a definujeme *Nové pravidlo formátování*. Styl  
formátování nechť je nastaven na možnost „Sady ikon“. Z nabídky  
Sady ikon pak vybereme *Tři hvězdičky*. Celá hvězdička se má zob-  
razit, pokud je číslo (sloupec *Typ* je třeba nastavit na volbu *číslo*)  
v buňce ostře větší než 0,5, půl hvězdičky se zobrazí, je-li hodnota  
menší či rovna 0,5, ale přitom je ostře větší než 0 (opět je sloupec  
*Typ* třeba nastavit na volbu *číslo*). Potvrzením této volby je celý  
postup dokončen.

## Literatura

- [1] *D. Huff*: How to Lie with Statistics (illustr. I. Geis), Norton, New York, 1954,  
český překlad dostupný online: <https://www.statistikaamy.cz/jak-lhat-se-statistikou/>.
- [2] Česko v číslech 2021. Český statistický úřad, Praha. Dostupné online:  
<https://csu.gov.cz/produkty/cesko-v-cislech-2021>.
- [3] *Mediaplanet*: Studie zkoumala 7 boosterů proti covidu, rozdíly mezi  
nimi jsou značné. 2022. <https://www.info-zdravi.cz/vakcinace-covid-19/studie-zkoumala-7-boosteru-proti-covidu-rozdily-mezi-nimi-jsou-znacne/>
- [4] *A. P S Munro, et al.*: Safety and immunogenicity of seven COVID-19 vac-  
cines as a third dose (booster) following two doses of ChAdOx1 nCov-19 or  
BNT162b2 in the UK (COV-BOOST): a blinded, multicentre, randomised,  
controlled, phase 2 trial. *The Lancet*, roč. 398(10318), (2021), s. 2258–2276,  
doi: 10.1016/S0140-6736(21)02717-3.