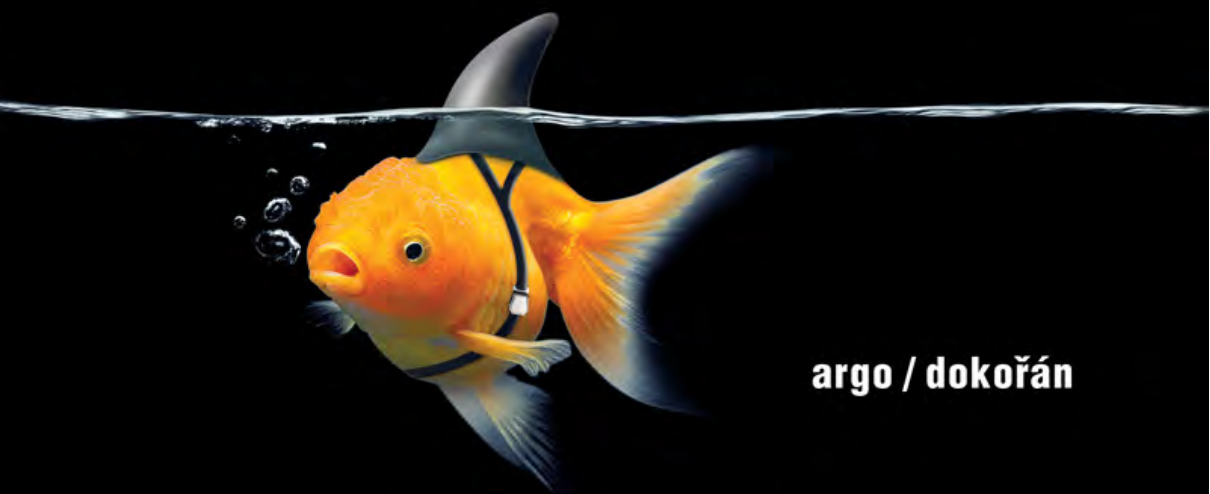


Tim Harford

JAK ČÍST ČÍSLA

**Deset pravidel pro orientaci
ve statistikách**



argo / dokořán

Tim Harford

JAK ČÍST ČÍSLA

**Deset pravidel pro orientaci
ve statistikách**

ARGO / DOKOŘÁN

Tim Harford

JAK ČÍST ČÍSLA

**Deset pravidel pro orientaci
ve statistikách**

Copyright © Tim Harford 2020

Translation © Jan Kalandra, 2022

Všechna práva vyhrazena. Žádná část této publikace nesmí být rozmnožována a rozšiřována jakýmkoli způsobem bez předchozího písemného svolení nakladatele.

Druhé vydání v českém jazyce (první elektronické).
Z anglického originálu *How to Make the World Add Up*.
Ten Rules for Thinking Differently About Numbers
přeložil Jan Kalandra.

Odpovědná redaktorka Klára Soukupová.

Obálka podle návrhu Pavla Růta, sazba a konverze
do elektronické verze Michal Puhač.

Vydalo v roce 2022 nakladatelství Dokořán, s. r. o., Holečkova 9, Praha 5,
dokoran@dokoran.cz, www.dokoran.cz,
jako svou 1157. publikaci (380. elektronická).

ISBN 978-80-7675-079-1

Učitelům na celém světě, těm mým zvlášť;
památce vzácného člověka Petera Sinclaira.

OBSAH

ÚVOD:	Prolhaná statistika	9
PRAVIDLO PRVNÍ:	Nahlédněte do své duše	23
PRAVIDLO DRUHÉ:	Zamyslete se nad svou zkušeností	43
PRAVIDLO TŘETÍ:	Pozor na předčasnou enumeraci	57
PRAVIDLO ČTVRTÉ:	Poodstupte si a pokochejte se výhledem	71
PRAVIDLO PÁTÉ:	Zjistěte si souvislosti	83
PRAVIDLO ŠESTÉ:	Ptejte se, na koho se zapomnělo	105
PRAVIDLO SEDMÉ:	Žádejte transparentnost, když počítač odmítá spolupracovat	119
PRAVIDLO OSMÉ:	Statistické podloží není samozřejmost	143
PRAVIDLO DEVÁTÉ:	Nezapomínejte, že i dezinformace můžou být krásné	163
PRAVIDLO DESÁTÉ:	Udržujte si otevřenou mysl	183
ZLATÉ PRAVIDLO:	Zajímejte se	201
	<i>Poděkování</i>	211
	<i>Poznámky</i>	213
	<i>Rejstřík</i>	237

PROLHANÁ STATISTIKA

Skutečný problém [...] nespočívá v tom, jak potvrdit nepravdivost tvrzení, ale naopak jak doložit pravost předmětu.

Umberto Eco¹

Určitě víte, že děti podle staré pověry nosí čáp. Ale víte taky, že je to pravda?

Můžu vám to statisticky dokázat.

Stačí porovnat odhadované populace čápů a roční počty nově narozených dětí v jednotlivých státech. Napříč Evropou mezi nimi panuje pozoruhodně silná shoda. Čím víc čápů, tím víc dětí – a naopak.

Souvztažnost je to napohled tak silná, že s přehledem zdolá tradiční překážku publikování v akademickém časopise. A skutečně už taky odborná práce s titulem „Děti nosí čáp ($p = 0,008$)“ vyšla. Bez zabíhání do podrobností k ní stačí poznamenat: všechny ty nuly naznačují, že nejde o náhodu.²

Možná už jste prokoukli, čím to. Velké evropské státy, jako je Německo, Polsko nebo Turecko, obývá hodně dětí i hodně čápů. V těch malých, například v Albánii nebo v Dánsku, je dětí i čápů málo. Údaje tedy sice vykazují jasnou shodu, ale z té nevyplývá, že by se děti rodily díky čápům.

Skoro jako by se statisticky dalo „dokázat“ cokoli – a tedy i pověra, že děti nosí čáp.

Přesně takový dojem byste si alespoň odnesli z četby knihy nazvané *Jak lhát se statistikou*. Tahle ohromně zábavná, ničeho se neštítící publikace z roku 1954, jejímž autorem byl tehdy málo známý americký novinář na volné noze Darrell Huff, si okamžitě vysloužila nadšenou kritiku v *New York Times* a postupem času se s více než půldruhým milionem prodaných výtisků jenom v angličtině stala asi nejčtenějším textem o statistice, který kdy vyšel.*

Knihy si všechnu tu oblibu i chválu zaslouží. Představuje zářný příklad statistické komunikace. A z Darrella Huffa udělala modlu všech nerdů. Epidemiolog Ben Goldacre, autor knižního bestselleru *Prolhaná věda*, se o „Huffanovi“ vyjádřil s obdivem a poukazem na to, že napsal „megadílo“. Za „poctu“ Huffově „klasice“ označil svou knihu *Naked Statistics* (Odhalená statistika) americký publicista

* Česky vyšlo jako *Jak lhát se statistikou* (přel. Martin Švehla, Praha: Brána 2013); zde používáme vlastní překlady z Huffova originálu *How to Lie with Statistics* (pozn. překl.).

Charles Wheelan. A uznávaný odborný časopis *Statistical Science* uspořádal Huffovi u příležitosti padesáti let od vydání jeho nejslavnějšího spisu retrospektivu.

I já jsem tu knížku zpočátku vnímal stejně. Coby náctiletého mě její četba ne skutečně bavila. Byla chytře napsaná, doprovázená hravými obrázkovými ilustracemi, umožnila mi nahlédnout za oponu manipulace se statistikami a ukázala mi, jak podvody s čísly probíhají, abych se příště už nenechal napálit.

Huff hýří příklady. Hned na začátku se zamýšlí nad tím, kolik si vydělá vysokoškolsky vzdělaný Američan s titulem ze slavné Yaleovy univerzity. Podle průzkumu z roku 1950 měli v té době absolventi z roku 1924 průměrný roční příjem blížící se v přepočtu dnešním 500 000 amerických dolarů. Takové hodnotě se ještě tak tak dá uvěřit – vždyť Yale patří mezi nejprestižnější vysoké školy v USA –, ale přesto: půl milionu dolarů je hodně peněz. To opravdu dosahuje průměr tak vysoko?

Ne. Huff vysvětluje, že tahle „podezřele opulentní“ cifra pochází z údajů poskytnutých samotnými respondenty, od nichž se však dá očekávat, že své příjmy z ješitnosti nadsazují. Průzkum navíc obsahuje údaje pouze o těch absolventech, kteří se dotazník obtěžovali vyplnit, a zároveň jen o těch, které se univerzitě podařilo dohledat. A kdo se hledá nejlépe? Pochopitelně ti bohatí a slavní. „Kdo jsou ty zaběhnuté ovečky, u nichž stojí v univerzitních análech poznámka ‚adresa neznámá?‘“ ptá se Huff. O milionářích, kteří ji vystudovali, si Yaleova univerzita záznamy určitě povede, ale někteří z těch, co absolvovali vedle nich, mohli sítlem její pozornosti snadno proklouznout. Ze všech uvedených důvodů plyne, že průzkum bude nutně vykreslovat značně zveličený obrázek.

Huff na dalších stránkách knihy prochází v rychlém sledu ohromnou šíří statistických zločinů od reklamy na zubní pastu opírající se o účelově vybrané výsledky výzkumu po mapy, které působí pokaždé jinak – podle toho, co v nich zvýrazníte. Jak Huff napsal: „Podvodníci už tyhle triky znají, takže je na poctivých občanech, aby se je v sebeobraně taky doučili.“

Četbou *Jak lhát se statistikou* si vybudujete obezřetnost vůči možnostem, jimiž vás čísla mohou klamat. Kniha je to chytrá i návodná.

Já se však už přes deset let snažím lidem statistické pojmy zprostředkovávat a číselná tvrzení ověřovat a za ty roky mě ta knížečka a všechno, co ztělesňuje, naplňují stále větším neklidem. Nevypovídá to koneckonců něco o statistice – i o nás samých, když nejúspěšnější kniha, která se tomuto tématu věnuje, představuje jedno velké varování před zkreslováním informací?

Darrell Huff vydal *Jak lhát se statistikou* v roce 1954. Jenže v témže roce se událo ještě něco: dva britští vědci, Richard Doll a Austin Bradford Hill, přišli s jednou z prvních studií, které přesvědčivě prokázovaly, že kouření cigaret způsobuje rakovinu plic.³

Bez statistiky by se toho závěru Doll s Hillem nedobrali. Výskyt rakoviny plic tehdy ve Velké Británii za pouhých patnáct let šestinásobně vzrostl a v roce 1950 už si Velká Británie v počtu pacientů s touto nemocí připsala světové prvenství, přičemž počet úmrtí na rakovinu plic poprvé předčil počty mrtvých s tuberkulózou. Statistický úhel pohledu byl nezbytný už jen k tomu, aby si člověk vůbec uvědomil, že se něco takového děje. Pohled jednotlivých lékařů na celou problematiku totiž nutně zůstával v rovině dojmologie.

A bez statistických údajů by nešlo ani doložit, že na vině jsou skutečně cigarety. Mezi lidmi kolovala hojně rozšířená domněnka, že za vzestupem rakoviny plic stojí automobilismus. Rozhodně totiž dávala smysl. V první polovině dvacátého století se automobily i s výfukovými plyny a podivně opojnou vůní asfaltu z nových silnic staly obvyklým jevem. Během téhož období vzrostl i počet případů rakoviny plic. Aby se člověk dobral pravdy - že tuto nemoc nezpůsobují auta, ale cigarety -, nestačilo se jen rozhlédnout. Bylo nutné, aby vědci začali pečlivě počítat a porovnávat. Neobešlo se to zkrátka bez statistik.

K hypotéze o cigaretách přistupovali mnozí s nedůvěrou, třebaže nepředstavovala žádnou novinku. Rozsáhlý výzkum, který měl přinést důkazy, že cigarety škodí zdraví, proběhl třeba už v nacistickém Německu - ostatně Adolfu Hitlerovi se kouření přičilo. A určitě ho potěšilo, když němečtí lékaři objevili, že cigarety rakovinu způsobují. Jenže sama skutečnost, že tabák „se protíví nacistům“, jeho oblibu pochopitelně nijak nesnížila.

A tak Doll s Hillem usoudili, že budou muset přijít s vlastním statistickým badáním. Richard Doll byl pohledný a tichý mladík, zdvořilost sama. Vrátil se z druhé světové války s hlavou plnou nápadů na to, jakou revoluci v lékařství může statistika vyvolat. Jeho mentor, Austin Bradford Hill, působil za první světové války jako pilot, než ho málem skolila tuberkulóza.* Hill měl charisma, pánilo mu to a provázela ho pověst nejvybranějšího lékařského statistika dvacátého století.⁴ A pátrání této dvojice v datech nakonec zachránilo řadu životů.

Jejich první studie o vlivu kouření na rakovinu byla zahájena 1. ledna 1948. Těžiště výzkumu, který vedl Richard Doll, tvořilo dvacet klinik v severozápadním Londýně. Vždy když byl do nemocnice přijat pacient s rakovinou, personál v tomtéž zařízení náhodně vybral dalšího nemocného stejného pohlaví a zhruba ve stejném věku. Pacienti s rakovinou i jejich protějšky byli dopodrobna dotazováni na to, kde bydlí, kde pracují, jaký život vedou, jak se stravují a jestli kouří. Týden co týden, měsíc co měsíc se políčka zaplňovala dalšími výsledky.

V říjnu roku 1949, ani ne dva roky od začátku studie, Doll přestal kouřit. Bylo mu sedmatřicet a kouřil od dospělosti. S Hillem právě zjistil, že silnému

* Později si připravil sladkou pomstu, když v práci považované za první řádně randomizovanou kontrolovanou studii demonstroval, jak tuberkulózu léčit.

kuřákovi nehrozí jen dvojnásobné, trojnásobné, nebo dokonce čtyřnásobné riziko, že onemocní rakovinou plic. Ta pravděpodobnost je vyšší šestnáctkrát.⁵

Výsledky Hill s Dollem zveřejnili v září roku 1950 a okamžitě se vrhli na ještě větší, dlouhodobější a dalekosáhlejší studii. Hill obepsal všechny britské lékaře - všech 59 600 - s prosbou o vyplnění dotazníku o jejich zdraví a kouření či nekouření. Oba vědci vycházeli z toho, že doktoři budou mít o tom, kolik toho vykouří, dobrý přehled. Zůstanou v rejstříku poskytovatelů zdravotních služeb, takže budou snadno dohledatelní. A když někdo z nich umře, dá se předpokládat, že bude stanovena kvalitní diagnóza příčiny jeho smrti. Naši dvojici tak stačilo jen čekat.

Na Hillovu výzvu zareagovalo přes 40 000 lékařů, byť ne všem se zamlouvala. Musíme si uvědomit, že kouření bylo tehdy mimořádně běžnou záležitostí, takže nikoho nepřekvapilo, když se zjistilo, že z prvotního výzkumného vzorku kouří 85 % doktorů mužského pohlaví. Jenže poslouchat, že se možná pomalu zabíjíte, zvláště když vám k tomu slouží vysoce návykový prostředek, by se asi nelíbilo nikomu.

Jeden doktor si Hilla odchytil na večírku v centru Londýna. „Tak to jste vy, ten maník, co chce, abychom přestali kouřit,“ spustil příkře.

„Ale kdepak,“ odvětil výzkumník, sám tehdy ještě kuřák dýmek. „Zajímá mě jen, jak umřete, jestli s kouřením nepřestanete. A stejně tak se chci přesvědčit, jak umřete, když se stanete nekuřákem. Volba, jestli s tím skončit, nebo ne, je na vás. Mně je to jedno. Vaši smrt si zaznamenám tak jako tak.“⁶

Nezapomněl jsem se zmínit, že původně vystudoval ekonomii? To odtud ten šarm.

Studie s lékaři pokračovala roky, ale výzkumné duo mělo brzy k dispozici dostatek dat, aby mohlo publikovat jasný závěr: kouření způsobuje rakovinu plic, a čím víc člověk kouří, tím víc mu onemocnění hrozí. A navíc - tady šlo o nové zjištění - kouření vede i k infarktu.

Doktoři si zbytečně nezahrávají. Když v roce 1954 výzkum vyšel přímo v jejich odborném časopise *British Medical Journal*, snadno si spočítali, co z něj pro ně plyne. Hill s kouřením skončil ještě téhož roku a stejně se zařídila i řada jeho kolegů. Lékaři se stali první jasně rozeznatelnou sociální skupinou ve Velké Británii, která od kouření ve velkém pouštěla.

V roce 1954 se tedy současně objevila dvě pojetí statistiky. Pro mnoho a mnoho čtenářů knihy *Jak lhát se statistikou* od Darrella Huffa představovala hru plnou šizůnků a nepoctivců, u které se dobře pobavili, když společně s autorem lsti těch podvodníků odkrývali. Jenže Austinu Bradfordu Hillovi a Richardu Dollovi statistika k smíchu nebyla. Při jejich hře bylo v sázce to nejcennější, co máme, a pokud se hrála poctivě a podle pravidel, uměla životy zachránit.

Vysoká cena, kterou kvalitní a rychle i poctivě získané statistiky mají, se zničehonic zcela jasně prokázala na jaře roku 2020. Svět bral útokem nový typ koronaviru. Politici museli bleskově přijímat rozhodnutí s takovými společenskými dopady, jaké nikdo celá desetiletí nezažil.

Tato přelomová rozhodnutí se opírala o pátrání epidemiologů, zdravotnických statistiků i ekonomů v datech. V ohrožení mohly být desítky milionů lidských životů, ale i miliardy živobytí. Zmiňovaní odborníci se usilovně snažili dopátrat toho, před čím svět stojí, jenže důkazy se jim hledaly zoufale pomalu. Epidemiolog John Ioannidis v polovině března 2020 napsal, že covid-19 „se může z hlediska důkazů stát fiaskem, jaké se neopakuje častěji než jednou za sto let“.⁷ O otázkách života a smrti se rozhodovalo prakticky naslepo.

Už za několik týdnů začali nicméně pátrači v datech načrtávat obrázek některých klíčových vlastností viru i nemoci, kterou způsobuje. Došlo nám, že ji řada lidí šíří, i když ještě sami nevykazují její příznaky. Poznali jsme, že u některých se příznaky vůbec neprojeví, a zásadní otázka zněla, kolik přesně jich je. (Dnes si myslíme, že asi čtvrtina, ale když se nad tím člověk na chvíli zamyslí, začne si uvědomovat, že bez pečlivého pátrání v datech se odpovědi dobereme jen těžko.) Brzy nám bylo jasné, že starší ročníky nemoc ohrožuje výrazně víc než ty mladší. Dokonce jsme měli k dispozici rozumný odhad její smrtnosti – v bohatých státech s velkým počtem starších obyvatel má činit zhruba jedno procento. O tomto údaji se zuřivě debatovalo, a s vývojem lékařských poznatků i mutací viru se také mění. Přesto počáteční odhady obstály překvapivě dobře.

Ačkoli některé záhady přetrvaly, obrázek nákazy nabýval postupem času – během nějž se nám dařilo navyšovat počet provedených testů, analyzovat případové studie, a dokonce sekvenovat mutující genetický kód viru – stále ostřejších obrysů. Tak velký důraz se na testování nekladl jen tak pro nic za nic – představovalo totiž možnost, jak si pomocí shromážděných dat vizualizovat skrytou hrozbu, zvlášť když se testy řádně utřídí a zanalyzují, jako to umí například britský Národní statistický úřad. Statistiky fungovaly jako pandemická obdoba radaru.

Tuto knihu jsem měl v hlavních obrysech napsanou dávno předtím, než se virem Sars-Cov-2 nakazil první člověk. Nijak mě nicméně neuspokojilo, když jsem sledoval, jak často smrtící pandemie potvrzuje tři hlavní argumenty, které už roky prosazuju.

Zprvė, naše emoce, předem utvořené názory a politická stanoviska nám dokážou krutě zdeformovat schopnost vykládat si důkazy. Tímto problémem, ústředním pro celé poselství knihy, se budeme zabývat v první kapitole.

Popisovaný jev v reálu ilustruje americká reakce na pandemii. Prezident USA Donald Trump koncem února 2020 prohlásil: „Však on ten virus zmizí. Jednoho dne se jako zázrakem vytratí.“ Zbožná přání mají svou moc. Ještě čtyři týdny poté

Trump vyjadřoval naději, že se na Velikonoce všichni zase sejdou v kostele.⁸ Když pak v zimě vyklízel Bílý dům, nemoc si už vyžádala životy 400 000 Američanů.

Trump je člověk, na kterého si většina lidí utvoří vyhraněný názor. Dalo by se čekat, že kdyby prohlásil, že zmrzlina člověku příjemní slunečný den, někteří Američané by od té doby nic jiného nejedli, zatímco jiní by před zmrzlinářstvími hlasitě protestovali. Totéž se odehrálo v případě covidu. Zvrácené, bezohledné odmítání si alespoň nasadit roušku se pro mnohé Trumpovy příznivce stalo odznakem hrdosti, a naopak jeho odpůrci zacházeli do opačného extrému – všiml jsem si výrazného tweetu, ve kterém jeden americký novinář vysvětloval, že se pandemie ve Velké Británii „vymkla kontrole“, protože lidé nenosí roušky při procházce v parku. Tohle tvrzení se z britského pohledu jevílo jako nehoráznost – vždyť důkazy hovoří o tom, že hrozba přenosu viru při procházce venku je minimální, rouška nerouška! A navíc v době, kdy se objevilo (koncem ledna 2021) se počty britských případů nákazy kontrole nevymykaly, naopak rychle klesaly.

Onen tweet se dal vyložit jedině jako výstřel v politicky vyostřené bitvě o zodpovědné nošení roušek, ve které ani jeden znesvářený kmen nezajímalo hledání pravdy. Kmenové uspořádání se dá paradoxně mnohem snáze odhalit, když si člověk zachová odstup. Kdo patří ke kmeni republikánů nebo demokratů, je do boje natolik vtažen, že se mu myšlení zastře. Lépe to zkrátka pálí těm, kteří náleží ke kmeni nechápajících nestraníků.

Zadruhé platí, že politická rozhodnutí mají určující vliv na to, jaké statistiky se budou shromažďovat a zveřejňovat – a jaké se budou přehlížet či zametat pod koberec. Třeba hned na začátku krize patrně zbrzdil volný tok otevřených statistik jeden geopolitický spor – a my se k tomuto typu problému ještě vrátíme v osmé kapitole. Tchaj-wan si opakovaně stěžoval, že koncem prosince 2019 Světové zdravotnické organizaci předal významné poznatky o přenosu viru mezi lidmi – nicméně WHO ještě v polovině ledna na Twitteru uklidňovala, že Čína žádné důkazy o přenosu mezi lidmi nemá. (Tchaj-wan není členem WHO, protože si jeho území nárokuje Čína, která po ostatních státech požaduje, aby jeho samostatnost neuznávaly. Je možné, že za údajně pomalým předáváním statistik stála právě tato překážka.)⁹

Hrálo to vůbec roli? Prakticky jistě – vzhledem k tomu, že se počet případů jednou za několik dní zdvojnásoboval, se už nikdy nedozvíme, co by se stalo, kdyby varování přišlo o pár týdnů dřív.

A konečně – statistiky nám ukazují jevy, které by pro nás jinak zůstaly nepozorovatelné. Neslouží jen k dozdobení novinového článku nebo jako zbraň v politické rozepři. Rozdíl mezi poctivými statistikami a chybnými či chybějícími daty může rozhodovat o životě a smrti.

Když jsem na jaře roku 2020 dopisoval první vydání *Jak číst čísla*, spoustu věcí jsme nevěděli a scházela nám spousta dat. Vzhledem k tomu, že počty mrtvých

tehdy exponenciálně rostly, však nebyl čas prostě si počkat, až budeme vědět víc. Vlády uložily ekonomiky k umělému spánku – za jediný týden na konci března se do evidence nezaměstnaných zapsaly více než tři miliony Američanů, pětkrát víc než o týden dřív. A následujících sedm dní bylo ještě horších: do statistik přibýlo dalších šest a půl milionu lidí bez práce. Opravdu byly možné dopady na zdraví obyvatelstva natolik katastrofální, aby se jimi dalo omluvit, že tolik lidí zničehonic přišlo o zdroj obživy? Jevilo se to tak – ale jen z odhadů, které epidemiologové na základě výrazně omezených informací mohli s nejlepším svědomím činit.

Těžko si představit výjimečnější ilustraci toho, jak moc za běžných okolností bereme přesně, systematicky shromažďované statistiky jako samozřejmost. Statistické údaje k mnoha důležitým problémům, které koronaviru předcházely, dlouhá léta pracně sbírali trpěliví statistikové a často je dávali zdarma k dispozici ke stažení celému světu. Jenže my už jsme z existence tak luxusních podkladů rozmazlení a ledabyle nad nimi máváme rukou, že prý to jsou stejně jen „lži, hnusné lži a statistiky“. Případ covidu-19 nám připomíná, do jak zoufalých situací se můžeme dostat, když ty údaje prostě k dispozici nemáme.

Darrell Huff proměnil statistiky v kouzelnický trik, který pobaví, ale vážně se brát nedá. Ve mně už nicméně dávno před koronavirem začalo klíčit podezření, že s takovým přístupem dnes nepochodíme. Pozapomněli jsme, že statistiky nám mohou pomoci pochopit svět. Ne snad že bychom měli pocit, že všechny statistické údaje lžou, ale když se v nich máme dobrat pravdy, netušíme, jak je číst. A tak věříme tomu, čemu věřit chceme (podrobněji si to ukážeme v příští kapitole), a na všechno ostatní aplikujeme Huffovu reakci: výsměch, pokrčení rameny, případně obojí najednou.

Tento statistický cynismus není jen smutný, ale přímo tragický. Jestliže se poddáme dojmům, že už není v naší moci poznat, kde je pravda, znamená to, že jsme zahodili jeden životně důležitý nástroj. Nástroj, který nám ukázal, že cigarety zabíjejí. Nástroj, který představuje naši jedinou reálnou možnost, jak najít cestu ven z koronakrizy – a obecněji jak porozumět složitému světu, v němž žijeme. Pokud však sklouzneme k tomu, že budeme automaticky zavrhnout libovolná statistická tvrzení, která se nám nepozdávají, nepomůže nám ani on. Pochopitelně bychom neměli naivně věřit všemu, co nám kdo nabulíkuj, nicméně protilátkou na naivitu není nihilismus, ale víra ve vlastní schopnost se zvědavostí i zdravou mírou nedůvěry informace posuzovat.

Navzdory tomu, co se mezi lidmi traduje, nejsou statistiky žádný podfuk, byť za kouzla je skutečně svým způsobem považovat můžeme. Nešíří kouřovou clonu, ale naopak odkrývají nové roviny. Kvalitní statistické údaje pro nás znamenají totéž co pro hvězdáře dalekohled, pro bakteriologa mikroskop

nebo pro radiologa rentgen. Pokud se jim otevřeme, odhalí ve světě kolem nás i v nás samých to, co bychom jinak přehlíželi - protože je to příliš maličké, nebo příliš obří.

Tuto knihu jsem napsal s jedním hlavním cílem, a to přesvědčit vás, abyste se na svět dívali očima Richarda Dolla a Austina Bradforda Hilla, ne téma Huffových. Chci vám dokázat, že se statistikou se nemusí lhát, ale dá se s ní osvětlit skutečnost tak, že ji vidíme nezkaleně a bez klamů. K tomu vám musím předvést, že statistické uvažování může být k užitku i vám samým, že si díky němu můžete vyhodnocovat výroky, jimiž vás obklopují média, sociální sítě i každodenní hovory. Chci vás to naučit, i když o statistice dosud třeba nic nevíte, a chci taky, což je neméně důležité, abyste si uměli sami zjistit, kde najdete pomoc, které můžete věřit.

Pozitivní na tom je, že vás to bude bavit. Když se doberete jádra statistického příběhu, zažijete skutečné uspokojení, posílíte si sebevědomí, ukojíte přitom svou zvědavost a nakonec odejdete s pocitem, že něco ovládáte. Že rozumíte, místo abyste se jen zpovzdálí ušklíbali. Přístup Darrella Huffa je jako nezdravé jídlo - napohled lákavé, ale po chvíli se člověku přejí. A škodí mu. Opakem statistických chipsů a hranolků však nejsou jen ovesné vločky a řepa, nýbrž chutný a pestrý jídelníček, který uspokojí hlad po vědě.

V této knize budu popisovat, co jsem se sám naučil od roku 2007, kdy jsem od BBC dostal nabídku uvádět rozhlasový pořad s názvem *More or Less* (Víceméně), který se zabývá čísly ve zpravodajství a v životě obecně. Odcházeli z něj totiž jeho původní autoři, novinář Michael Blastland a ekonom sir Andrew Dilnot. Měl jsem pro tu roli menší kvalifikaci, než si asi BBC představovala, nevystudoval jsem totiž statistiku, ale teoretickou ekonomii. Je pravda, že jsem se díky tomu v číslech trochu vyznal, byť spíš pasivně - uměl jsem poznat, že je něco špatně nebo to zavání nekalostí, ale na o moc víc to nestačilo.

Právě tehdy jsem se nicméně začal odklánět od názorů Darrella Huffa.

Týden co týden jsme s kolegy hodnotili statistické výroky, které vypustili z úst politici nebo které se palcovým písmem skvěly v novinových titulcích. Často nějakým způsobem překrucovaly pravdu, nicméně samotné ověření údajů se nám nikdy nezdálo jako uspokojivá reakce. A tak jsme postupně zjišťovali, že za každým tím tvrzením - pravdivým, nepravdivým, či na pomezí mezi nimi - se skrývá ohromující svět, který čeká na probádání a vysvětlení. Ať už jsme zkoumali výskyt mrtvice, důkazy, že dluhy brzdí růst hospodářství, anebo dokonce to, kolikrát se v *Hobitovi* vyskytuje zájmeno „ona“, dokázala čísla svět osvětlit, ale i zatemnit.

Jak tolik neúprosně dokládá epidemie koronaviru, k utváření svých rozhodnutí potřebujeme spolehlivá čísla - my jako jednotlivci, jako různé organizace

i jako celá společnost. A právě jako v případě koronaviru v dějinách často platilo, že se statistiky sbírají jen tehdy, když stojíme tváří v tvář krizi. Jako příklad postačí uvést míru nezaměstnanosti – údaj o tom, kolik lidí hledá práci a nemůže ji najít. Ta dnes z pohledu vlád, které chtějí porozumět stavu ekonomiky své země, představuje jeden ze základních ukazatelů, nicméně ještě ve 20. letech 20. století vám to, kolik lidí je bez práce, nebyl nikdo schopný říct.¹⁰ Údaje, které tuto otázku uměly zodpovědět, začaly státy sbírat, až když se vlivem hlubokých hospodářských propadů stala politicky relevantnější.

Náš velký, matoucí svět se takovými otázkami, na něž nám dovede dát odpověď jen bedlivá pozornost věnovaná číslům, jen hemží. Přináší nám Facebook spíše štěstí, nebo smutek a umíme vysvětlit, proč se na něm různí lidé chovají různě? Kolika živočišným a rostlinným druhům hrozí vyhynutí, jaké procento z celku tvoří a je příčinou tohoto stavu změna klimatu, zintenzivňování naší zemědělské činnosti, nebo třeba něco úplně jiného? Zrychluje, či zpomaluje tempo lidských inovací? Jak vážné dopady má drogová krize na zdraví obyvatel Střední Ameriky? Pijí dnes dospívající méně než předchozí generace – a pokud ano, čím je to dáno?

Když nás fanoušci *More or Less* chválili za to, jak „vyvrácíme nepravdivé statistiky“, čím dál víc mě to zneklidňovalo. Ano, dělali jsme to a bavilo nás to. Jak jsem si však přitom postupně byl stále jistější v kramflecích, začalo mi docházet, že opravdová radost netkví v odhalování lží, ale ve snaze pochopit pravdu.

Při práci na *More or Less* jsem zjistil, že si člověk při pátrání v datech do překvapivě velké míry vystačí s principy zdravého rozumu. A právě ty hodlám v téhle knize shrnout. Většina našeho týmu badatelů a producentů stejně jako já při zacházení s čísly postrádala formální vzdělání. Přesto nám i ve vysoce odborných záležitostech často stačilo několik jednoduchých otázek – a možná jedno dvě vyhledávání na internetu –, abychom dospěli k opravdu uspokojivým odpovědím. Samozřejmě že někdy by se nám doktorát ze statistiky hodil, ale na to, abychom se správně zeptali, jsme ho nikdy nepotřebovali. A nepotřebujete ho ani vy.

Těsně před Vánoci roku 1953 se do newyorského hotelu Plaza sjeli čelní představitelé tabákových společností. Velká studie R. Dolla a A. B. Hilla měla sice vyjít až v příštím roce, ale výrobci cigaret už věděli, že se pro ně ohledně vědeckých poznatků začíná situace vyvíjet dost špatně. A tak si dali dostaveníčko, aby vymysleli, jak na blížící se krizi zareagovat.

Jejich reakce byla naneštěstí takřka geniální a od té doby slouží jako propagandistický standard.

Začali totiž kalit vodu. Začali stávající výzkumy zpochybňovat, volali po dalším zkoumání, financovali výzkum jiných problémů – například syndromu nezdravých budov či nemoci šílených krav – a doufali, že by se jim mohlo podařit

přesvědčit média, aby se pro ně nadchla. Rozjeli velkovýrobu pochybností.¹¹ Motto „naším produktem jsou pochyby“ později zasvěceným připomínal i tajný interní dokument jedné tabákové firmy.¹²

Není divu, že když se bavíme o přesvědčování, míváme nezdárka na mysli, že někdo někoho přiměje, aby věřil něčemu, čemu by věřit neměl – a my si tento jev rozebereme v příští kapitole. Jenže někdy nejde o to, že bychom příliš ochotně přijímali něčí stanovisko, nýbrž o to, že si hledáme důvody, abychom nemuseli věřit vůbec ničemu. Kuřáci měli kouření rádi, byli na nikotinu fyzicky závislí a chtěli pokud možno kouřit dál. Situace, kdy si milovníci cigaret budou s pokrčenými rameny říkat, že se ve všech těch matoucích vyjádřeních prostě nevyznají, tabákovému průmyslu vyhovovala. Jeho cílem tak nebylo přesvědčit své zákazníky, že jeho výrobky jejich zdraví neškodí, ale zahalit pochybnostmi statistické důkazy o tom, že je ohrožují na životě.

Pochybnosti jdou přitom vytvářet opravdu snadno. Psycholožka Kari Edwardsová a její kolega Edward Smith provedli před dvěma desítkami let pokus, v němž požádali Američany, aby přicházeli s argumenty pro a proti k aktuálním politicky citlivým otázkám, například právu na umělé přerušování těhotenství, užívání fyzických trestů při výchově dětí, adopcí dětí homosexuálními páry, kvótám pro zaměstnávání menšin či trestu smrti pro mladší šestnácti let.¹³ Asi vás nepřekvapí, že pokusné osoby trpěly předpojatostí – příliš se jim nedařilo vymýšlet takové argumenty, jimiž by jejich protivníci své názory obhajovali. Pozoruhodnější nicméně bylo, že se podle zjištění Edwardsově se Smithem tyto předsudky zřetelněji vyskytovaly v případě argumentů proti. Nedůvěra se šířila snáze než vlastní přesvědčení. Účastníkům experimentu se mnohem snáze argumentovalo proti názorům, s nimiž nesouhlasili, než na podporu těch, které sami zastávali. Ukázalo se, že pochybnosti mají zvláštní moc.

Pochybnosti jdou taky snadno prodávat, protože tvoří součást procesu vědeckého poznávání a vědecké debaty. Zpochybňovat důkazy se většina z nás učí – nebo by alespoň měla – ve škole. Heslo jednoho z nejstarších vědeckých spolků, Královské společnosti, zní *nullius in verba* – „na ničí slova nepřisahej“ nebo taky „nedůvěřuj, prověřuj“. Když bude nějaká lobbistická skupina chtít popřít statistické důkazy, vždycky bude moct poukázat na určitou stránku současného stavu vědy, která není dořešená, přidat poznámku, že jde o nesmírně komplikovanou záležitost, a vyzvat k dalšímu zkoumání. Takové výroky navíc budou působit vědecky, skoro až prozíravě. Budou však budit falešný dojem: že nikdo nic neví.

Metody tabákového průmyslu mezitím našly široké uplatnění.¹⁴ Nejzřetelněji je dnes využívají popírači změny klimatu, nicméně už také překročily hranici vědy a vtrhly do politiky. Historik Robert Proctor, který tabákový průmysl zkoumá celá desetiletí, označuje moderní politiku za „zlatý věk nevědomosti“.

Stejně jako chce řada kuřáků kouřit i nadále, i my v politických otázkách často vehementně lpíme na názoru, který jsme si utvořili jaksi podvědomě. Politikům tak stačí jen zpochybňovat případné důkazy, které by naše stanovisko mohly nahлотat.

Nechvalně známý je koneckonců výrok Steva Bannona, pravé ruky Donalda Trumpa, který spisovateli Michaelu Lewisovi přiznal: „Demokrati nám jsou putna. Opravdová opozice jsou média. A ty je potřeba zaplavit sračkama.“¹⁵

Poučný je v této souvislosti i příběh dalšího pojmu spojovaného s Donaldem Trumpem, „fake news“. Jako „lživé zprávy“ se zprvu označoval jeden velice konkrétní jev: když na různých webech vycházely smyšlené články s cílem získat co nejvíc prokliků ze sociálních sítí, a tím pádem i větší příjmy z reklamy. Jako už legendární příklad poslouží tvrzení, že Trumpovu prezidentskou kandidaturu podpořil papež. Když Trump vyhrál, na chvíli zavládla morální panika, v níž si seriózní komentátoři dělali obavy, že k odevzdání hlasu Trumpovi přispěli důvěřiví voliči, které k tomu svedly tyto nehorázné lži, jimž uvěřili.

Tato panika byl jeden velký omyl. Akademické studie zjistily, že fake news nikdy nebyly příliš rozšířené či vlivné, protože většina z nich dopadala na malý počet vysoce konzervativních starších voličů, kteří by Trumpa nejspíš volili i tak. Tyto nepodložené zprávy navíc brzy začaly způsobovat mnohem menší problémy, protože se sociální sítě probraly a jejich hrozbě začaly čelit.¹⁶

Mocnou se nicméně stala *myšlenka* „fake news“ – a začala fungovat jako výmluva, proč zavrhnout libovolné nepohodlné tvrzení nehledě na zdroj, z něž pochází; jako moderní obdoba cynického aforismu, že „existují tři druhy lži: lži, hnusné lži a statistiky“. Donald Trump se svým zvráceným nadáním přetvořit složitý problém ve zbraň na svoje odpůrce začal tento termín používat k demoinizaci běžných novinářů. A stejně se chovala i řada dalších politiků včetně tehdejší britské premiérky Theresy Mayové a jejího opozičního protějšku, předsedy labouristů Jeremyho Corbyna.

Fake news měly ohlas, protože vycházely z jedné nešťastné pravdy: jak se ještě přesvědčíme, nedbalých novinářů se dokonce i v hlavních médiích vyskytuje spousta. A ti seriózní a zodpovědní, kteří svoje výroky pečlivě opatřují zdroji, jsou pak s hokynáři, prodávajícími historky jako o papežově podpoře Trumpovi, házení do jednoho pytle.

Skutečnost, že se zdravá nedůvěra může snadno „zdrccnout“ do podoby jedovatého cynismu, nám opět připomněl koronavirus. Během období prvních lockdownů, na jaře 2020, přišlo hnutí „skeptiků“ s otázkami, kdy by mohly škody páchané omezováním společnosti začít převažovat nad výhodami plynoucími ze zpomalení šíření nákazy – a odstartovalo tak obtížnou, ale podstatnou debatu. Do léta, kdy se výskyt viru scvrkl na minimum, se nicméně řada nejvýraznějších hlasů jala hlasitě proklamovat, že pandemie skončila, a přidávat

prognózy, že žádná druhá vlna nebude. A když se s podzimním raketovým nárůstem počtu mrtvých ukázalo, jak trapný omyl taková tvrzení představovala, jejich autoři se běžně uchylovali ke kalení vody výroky, že spoustu chyb napáchali i mainstreamoví odborníci, takže ani jim se nedá věřit.

Čelní postavení mezi těmito údajnými chybami zaujalo údajně výrazné nadhodnocení smrtnosti viru, kterého se v březnu 2020 měla dopustit Světová zdravotnická organizace (WHO), když prý tvrdila, že činí 3,4 %, což je několiknásobně víc než jeho skutečná smrtnost. Chyba lávky. Výrok WHO byl překroucený – došlo k záměně jedné definice za jinou, napohled podobnou. (Rozostřeným definicím se budeme podrobněji věnovat ve třetí kapitole.) WHO spočítala procento úmrtí z oficiálního počtu všech případů – 3,4 % –, které se však nerovnal smrtnosti viru z toho prostého důvodu, že v březnu 2020 prošly testováním pouze osoby s vážnými příznaky.¹⁷

Vidíte tu podobnost? Děj se zde odvíjí podle téhož scénáře, jaký v 50. letech použily tabákové společnosti: nadhodit, že odborníci ve skutečnosti nerozumějí tomu, k čemu se vyjadřují, a zanechat lidi v nečinnosti vyvolané pochybami. U mnohých z těch, kteří se sami pasovali do role „odmítačů“ uzávěr, se nakonec ukázalo, že se nesnaží dosáhnout hlubšího pochopení, ale že jde o profesionální vyvolávače kontroverzí, kteří z pozornosti pramenící z rozčilení chtějí vytěžit peníze nebo moc. Takovým lidem nemohlo nic posloužit lépe než publikum, které se zoufale a znechuceně vzdává vlastního názoru. Takoví lidé se zuřivě snaží, abyste se k důkazům, k odborníkům a k pečlivě ověřeným zprávám médií obraceli zády. Nejvíc je potěší, když si řeknete: „Dneska se už nedá věřit ničemu a nikomu.“ Vzpomeňte si na strategii Steva Bannona: zaplavit sračkama.

Svět, ve kterém spousta lidí uvěří všemu, mi dělá starosti. Mnohem víc se ale bojím takového, ve kterém lidé kromě vlastních ustálených názorů nevěří ničemu.

Na jaře roku 1965 zvažoval výbor amerického Senátu otázku života a smrti – zda na krabičkách cigaret bude muset být varování před jejich vlivem na zdraví. Jednomu odbornému znalci se vědecké důkazy nezdály příliš přesvědčivé, a tak si vzal na pomoc motiv čápů a dětí. Vysvětloval, že mezi počtem narozených dětí a počtem čápů v okolí panuje kladná korelace.¹⁸ Ta stará pověra, že děti nosí čáp, není pravdivá, pokračoval – jasně že není. Korelace není kauzalita. Čápi děti nenosí. Ale na větším území je jak pro děti, tak pro čápy víc místa. A podobně pouhá skutečnost, že kouření koreluje s rakovinou plic, v žádném případě neznamená, že kouření rakovinu *způsobuje*.

„Opravdu si myslíte, že u statistik o spojitosti kouření a nemocí panuje stejný příčinný vztah jako u čápů a dětí?“ zeptal se ho předseda výboru. Odborný znalec odpověděl, že mu „obojí připadá stejné“.¹⁹

Jmenoval se Darrell Huff.

Tabáková lobby mu zaplatila, aby udělal to, co mu šlo nejlíp: smíchal dohromady vtípné příklady s trochou statistických zajímavostí a jistou dávkou cynismu, a pomocí této směsi zahalil představu, že cigarety škodí, mračnem pochybností. Dokonce měl rozpracované pokračování svého nejslavnějšího díla, byť nakonec nevyšlo. Neslo název *Jak lhát se statistikou o kouření*.²⁰

Pochyby jsou mocná zbraň a statistiky snadný cíl. A tenhle cíl potřebuje obránce. Ano, se statistikou se dá lhát snadno – ale bez ní ještě snáz.*

Podstatnější však je, že bez statistiky nejde poznat pravda – pochopit svět, abychom ho mohli jako Richard Doll s Austinem Bradfordem Hillem změnit k lepšímu. K tomu, čeho tihle dva dosáhli, bylo zapotřebí jisté odbornosti i odhodlání, ale určitě si to nevyžadovalo genialitu ani vysokou matematiku. Spočítali prostě to, na čem záleželo – kuřáky, nekuřáky, případy rakoviny plic, případy srdečních chorob. Počítali je metodicky, trpělivě a z důkazů, které nahromadili, pečlivě vyvozovali závěry. Ty za ta léta prodloužily život desítkám milionů lidí, je samotné nevyjímaje: poté, co Hill přestal kouřit dýmku a zařadil se k Dollovi mezi nekuřáky, se oba dožili devadesátky.

Když spolu se statistikou použijeme i víru ve vlastní schopnosti a moudrost, začneme si všimnout tendencí, které by pro nás jinak zůstaly neviditelné. Moderní svět je hrozně velký, hrozně složitý a hrozně, hrozně zajímavý. Žije na něm skoro osm miliard lidí. V naší ekonomice přecházejí denně z ruky do ruky biliony dolarů. Průměrný lidský mozek má 86 miliard neuronů.²¹ Na internetu je okolo dvou miliard webů. A nový virus se může rozšířit z jediného člověka na tisíce, miliony – ba miliardy jiných. Ať se budeme snažit na světě, na druhých i na sobě pochopit cokoli, bez statistiky daleko nedojdeme – rozhodně ne dál než při snaze podívat se do kostí bez rentgenu, na bakterie bez mikroskopu nebo na nebe bez hvězdářského dalekohledu.

O Galileovu dalekohledu se traduje oblíbená historka: ještě ve chvíli, kdy otec astronomie čelil obvinění z kacířství, vysoce postavení římskokatoličtí kardinálové do přístroje, který sestavil, nechtěli ani nahlédnout a jen opakovali, že jde o kouzelnický trik. Že prý Galileo na Měsíci viděl hory? To má určitě jen špinavý objektiv. Že pozoroval měsíce Jupitera? To tak! Ty musejí být v tom teleskopu někde schované. Prostě se odmítali podívat.

O čtyři staletí později se nám té anekdotě – která je za ty roky mimochodem už trochu poupravená – hezky směje.²² Ale neměli bychom podléhat sebeuspokojení. Mnozí z nás se odmítají dívat na statistické důkazy, protože se bojí, že na ně někdo šije boudu. Připadáme si protřele, když bereme za svůj Huffův přístup a všechny statistiky cynicky zavrhneme. Žádní světáci však nejsme. To

* Statistikové mají tento aforismus v oblíbě. Často vidám, že se připisuje skvělému statistikovi Fredericku Mostellerovi, nicméně potvrdit toto tvrzení se mi nepodařilo.

se jen poddáváme populistům a propagandistům, kteří po nás chtějí, abychom krčili rameny, dali sbohem logice a důkazům a stáhli se do názorové ulity, ve které nám je dobře.

Já nicméně chci, abychom dělali něco jiného. Chci vám dát sebedůvěru, abyste se nebáli zvednout statistický dalekohled a jeho hledáčkem propátrali svět. Chci vám pomoci pochopit logiku, která se skrývá za statistickými pravdami, a uniknout ze sevření logických omylů, pocitů a kognitivních zkreslení, které dávají vzniknout lžím.

Podívejte se tedy do statistického teleskopu a rozhlédněte se. Ohromí vás, jak jasně uvidíte.