MYSLETE JASNĚ

Jak lépe rozumět světu kolem vás



MEZINÁRODNÍ BESTSELLER

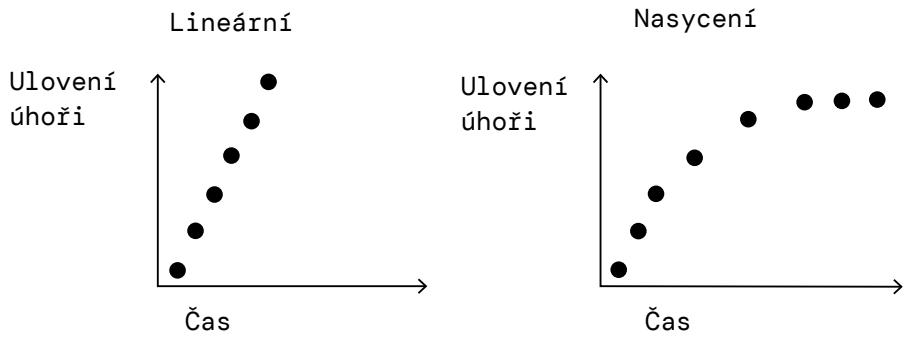
KIKO LLANERAS

DOPROVODNÉ MATERIÁLY K AUDIOKNIZE

progresguru

Pravidlo první

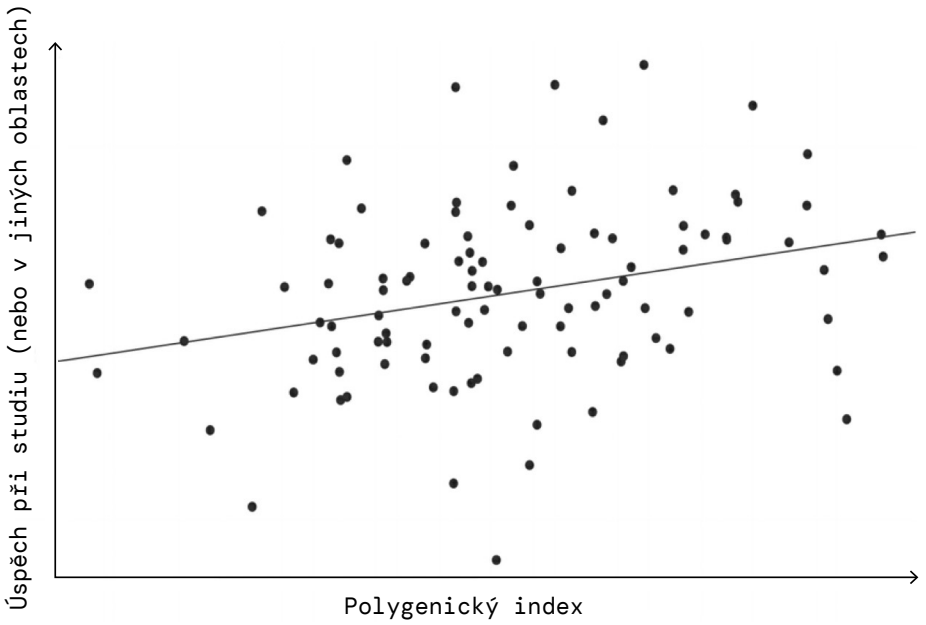
Žijete ve složitém světě



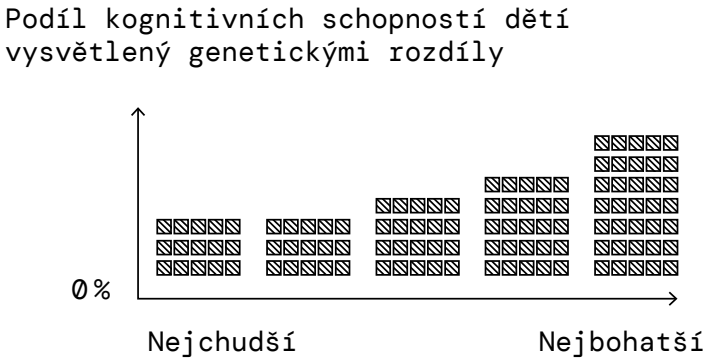
Obr. 1 Lineární vztah a nasycení. Kdyby byl vztah lineární, počet ulovených úhořů by rostl stejným tempem jako čas strávený lovem. Úhořů v řece ale ubývá, a proto úlovek postupně přestává růst.

Pravidlo první

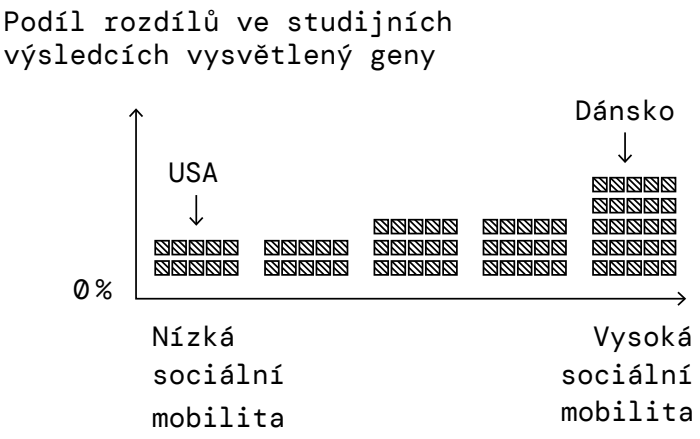
Vše má mnoho příčin



Obr. 2 Hypotetický polygenický index, který zachycuje 10 % rozdílů ve výsledcích v průběhu života, například ve studijním úspěchu nebo v příjmech v dospělosti. Každý bod je jeden člověk. Svislá osa ukazuje úspěch při studiu, vodorovná polygenický index.



Obr. 3 Podíl kognitivních schopností dětí vysvětlený genetickými rozdíly. U dětí z nejchudších rodin je menší než u dětí z nejbohatších rodin.



Obr. 4 Podíl rozdílů ve studijních výsledcích vysvětlený geny. V USA s nízkou sociální mobilitou je menší než v Dánsku, kde je sociální mobilita vysoká.

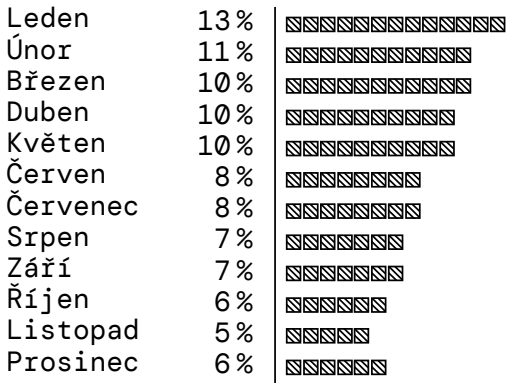
Pravidlo první

Složitost vzniká z jednoduchosti

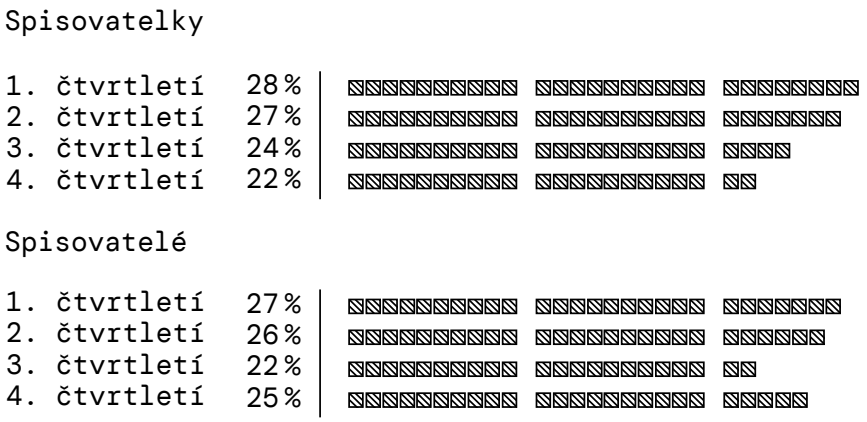


Obr. 5 Vlnění hejna špačků nad polem poblíž Kiryat Gat na jihu Izraele. Foto: Reuters, © Amir Cohen.

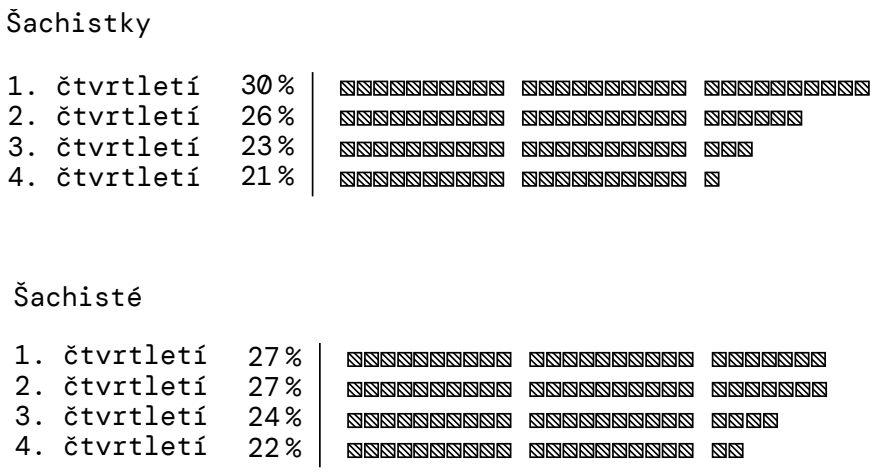
Některé věci jsou samy sobě příčinou



Obr. 6 Španělští fotbalisté v první a druhé lize a v hlavních evropských soutěžích v sezóně 2020/21 podle měsíce narození. Nejvíc se jich narodilo v lednu (13 %), nejméně v listopadu (5 %). Procenta jsou zaokrouhlená. Zdroj: Driblab.



Obr. 7 Španělské spisovatelky a spisovatelé na Wikipedii podle čtvrtletí narození. U spisovatelek je efekt výraznější: v prvním čtvrtletí se jich narodilo 28 %, ve čtvrtém 22 %.

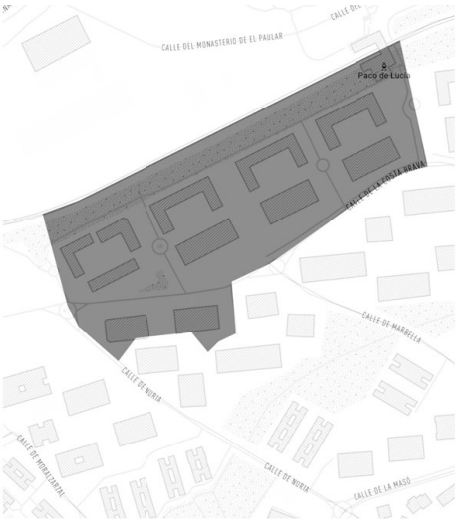


Obr. 8 Šachistky a šachisté na Wikipedii podle čtvrtletí narození. V prvním čtvrtletí se narodilo 30 % šachistek, ve čtvrtém jen 21 %.

Vše měřte



Stanice metra Bilbao (centrum)



Stanice metra Paco de Lucía (sever)

Obr. 9 Dvě madridské oblasti se stejným průměrným příjmem na osobu, zhruba 24 000 eur čistého ročně. Tribunal leží v centru u stanice metra Bilbao, Paco de Lucía na severním okraji města.

	Tribunal	Paco de Lucía
Průměrný příjem na osobu	24 000 eur	24 000 eur
Průměrný příjem domácnosti	49 000 eur	81 000 eur

Obr. 10 Tribunal a Paco de Lucía mají stejný průměrný příjem na osobu. Průměrný příjem domácnosti je ale v Paco de Lucía výrazně vyšší, protože tam žijí větší rodiny s dětmi.

	Tribunal	Paco de Lucía
Průměrný příjem na ekvivalentní osobu	21 000 eur	33 000 eur

Obr. 11 Průměrný příjem na ekvivalentní osobu je 21 000 eur v Tribunalu a 33 000 eur v Paco de Lucía. Podle tohoto ukazatele je Paco de Lucía bohatší.

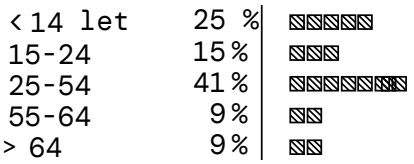
Pravidlo druhé

Zkoumejte svá data

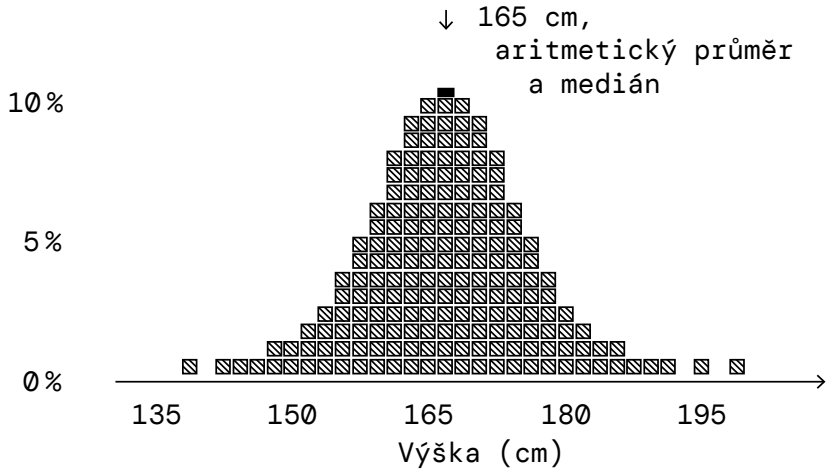
Sourozenec:	1	2	3	4	5	6	7
Majetek:	0	0	0	0	0	0	70 milionů

↑
Medián = 0

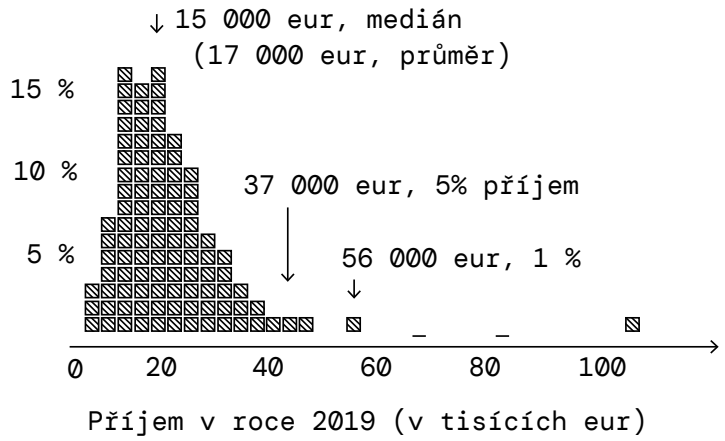
Obr. 12 Majetek sedmi sourozenců, z nichž jeden má 70 milionů eur. Průměr vychází na 10 milionů eur, medián je 0.



Obr. 13 Věkové rozložení světové populace. Největší skupinu tvoří lidé ve věku 25 až 54 let (41 %). Zhruba každý čtvrtý člověk na světě je dítě.



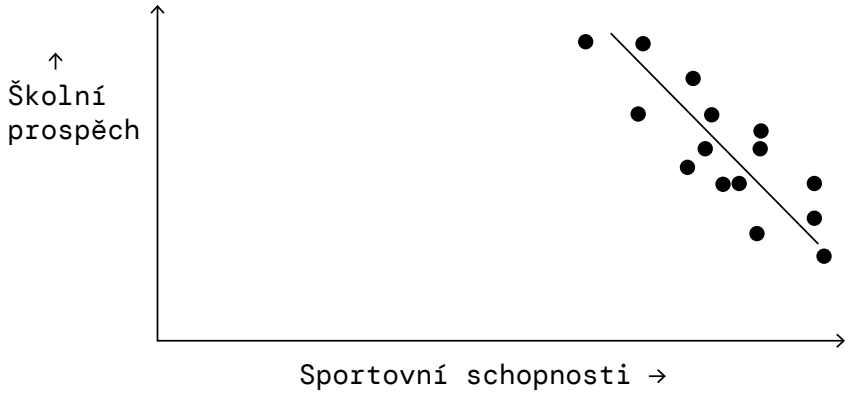
Obr. 14 Výška žen má normální neboli zvonovité rozložení. Průměr i medián je 165 cm a 68 % žen měří mezi 158 a 172 cm.



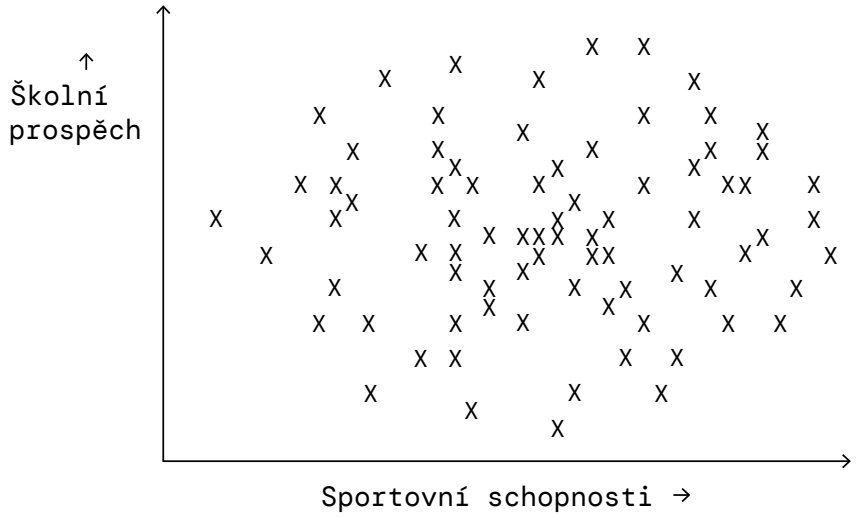
Obr. 15 Rozložení čistého příjmu na ekvivalentní osobu ve Španělsku v roce 2019. Je asymetrické: hodně lidí má nízké příjmy a s rostoucím příjmem jich ubývá. Medián je 15 000 eur, průměr 17 000 eur. Zdroj: INE.

Pravidlo třetí

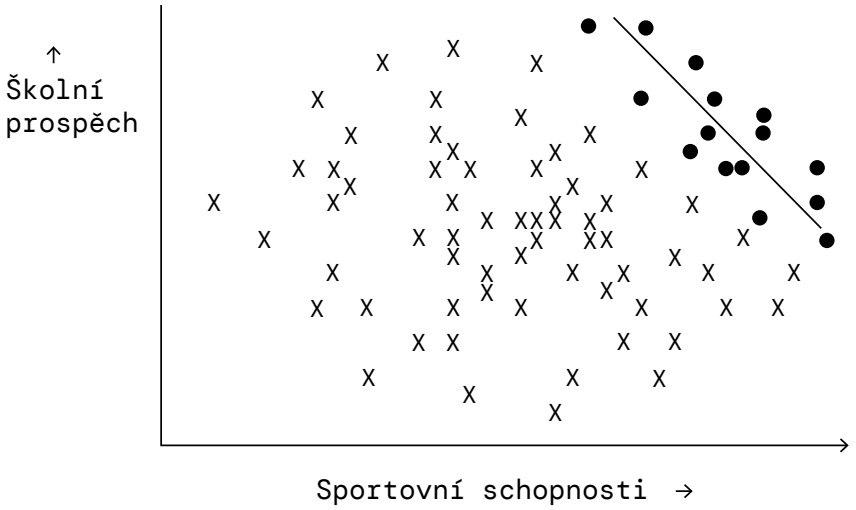
Pozor na zkreslení výběrem



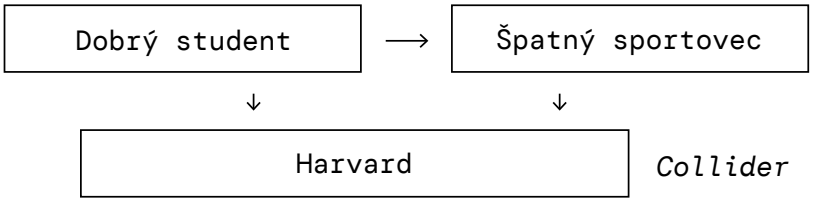
Obr. 16 Studenti Harvardu: čím lepší školní prospěch, tím horší sportovní schopnosti. Tento vztah ale vzniká jen ve vzorku, který je zkreslený výběrem.



Obr. 17 Akademické a sportovní schopnosti 70 libovolných mladých lidí. Obě schopnosti jsou na sobě nezávislé, a proto v grafu vidíme jen shluk bodů bez jakéhokoli vztahu.



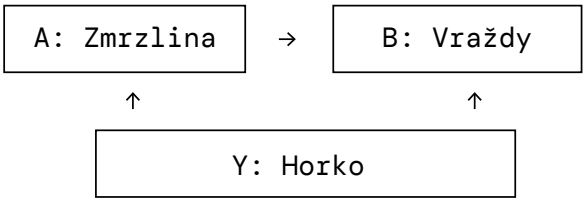
Obr. 18 Tentýž shluk se zvýrazněnými studenty Harvardu (černé body). Harvard vybírá studenty z horních 20 % podle studijních nebo sportovních výsledků. Mezi přijatými proto vzniká negativní vztah, který v celé populaci neexistuje.



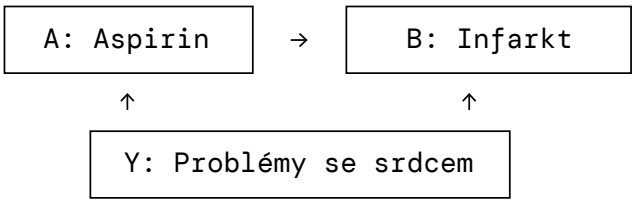
Obr. 19 Kauzální diagram zkreslení výběrem. Být dobrý student a být špatný sportovec má společný následek: dostat se na Harvard. Takové proměnné se říká collider.

Pravidlo čtvrté

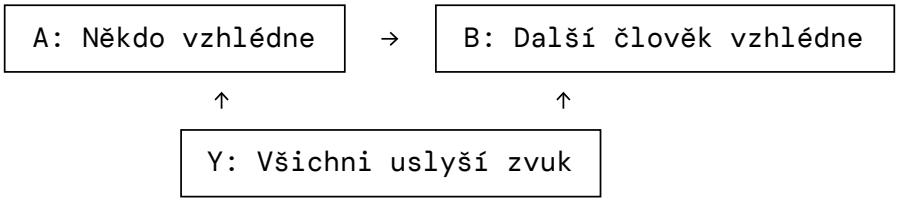
Korelace neznamená kauzalitu



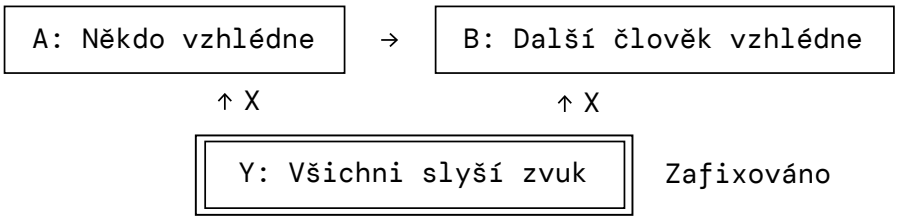
Obr. 20 Souvislost mezi prodejem zmrzliny (A) a počtem vražd (B) může vytvářet jejich společná příčina, horko (Y).



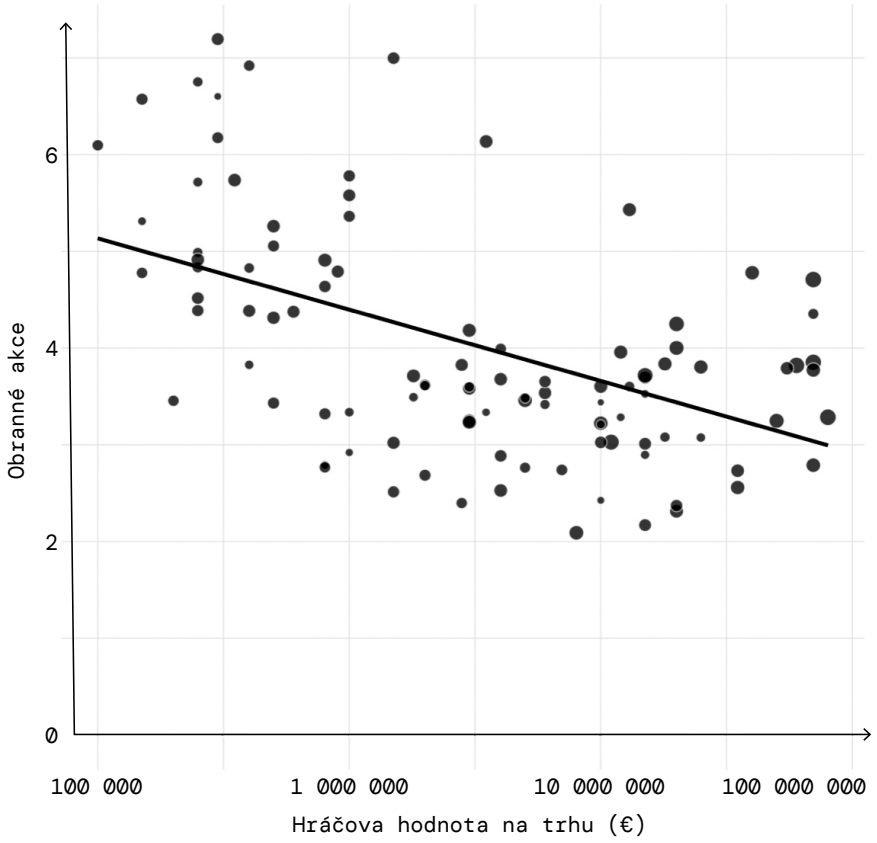
Obr. 21 Problémy se srdcem (Y) zvyšují riziko infarktu (B) i pravděpodobnost, že pacient dostane předepsaný aspirin (A). Účinek léku je proto zkreslený.



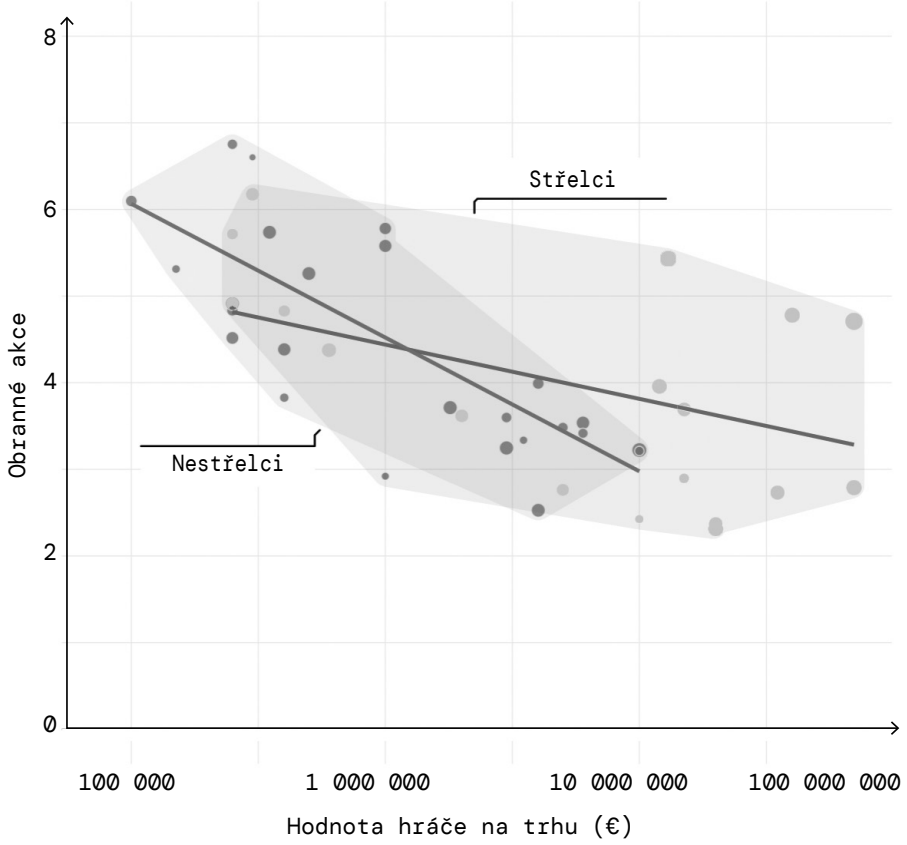
Obr. 22 Že se jeden chodec podívá nahoru (A) a po něm i další (B), může mít společnou příčinu: oba uslyší zvuk z oblohy (Y).



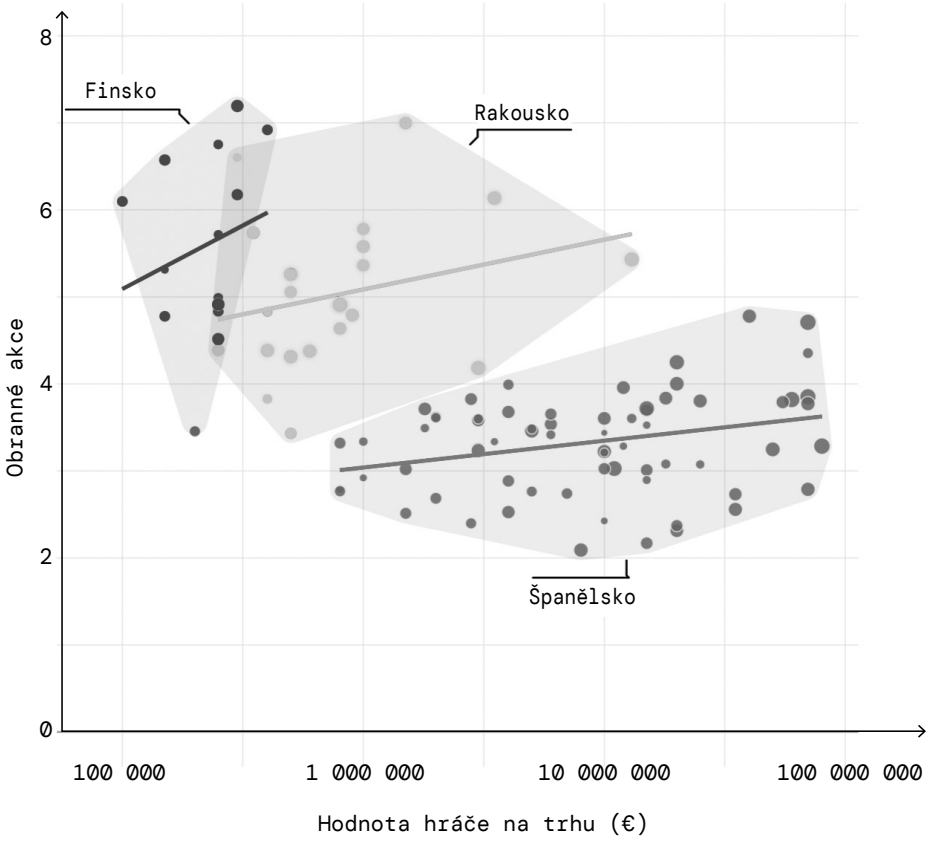
Obr. 23 Tentýž diagram po zafixování matoucího faktoru. Když pozorujeme jen situace bez zvuku, zbývá jediný zdroj souvislosti: A způsobuje B.



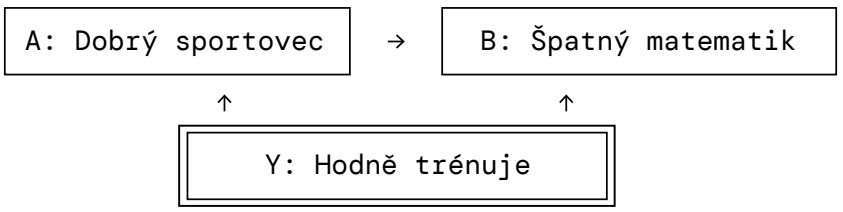
Obr. 24 Záložníci z různých evropských lig. Každý bod je jeden hráč: jeho tržní hodnota podle webu Transfermarkt a počet obranných akcí za uplynulý rok podle Driblab. Nejlepší obránci mají nižší tržní hodnotu.



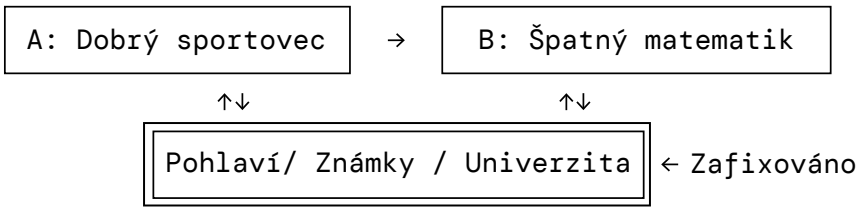
Obr. 25 Titíž záložníci rozdělení na střelce a nestřelce. Negativní vztah nezmizí: i mezi nestřelci jsou dražší hráči horší v obraně.



Obr. 26 Titíž záložníci rozdělení podle ligy: Finsko, Rakousko a Španělsko. V rámci jedné ligy jsou nejlepší obránci nejdražší. Liga je matoucí faktor a tento případ se označuje jako Simpsonův paradox.



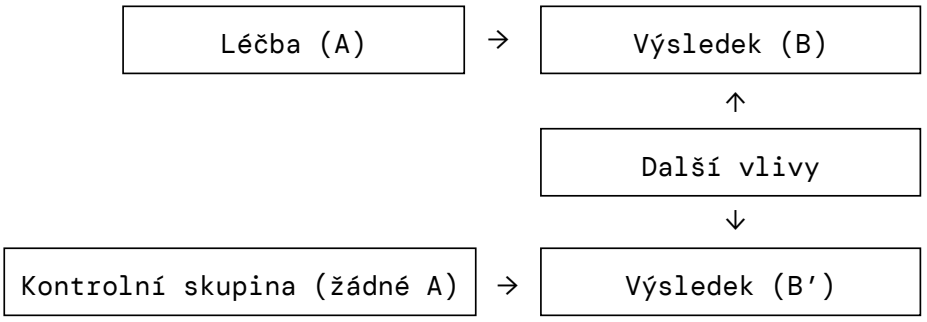
Obr. 27 Hodně trénovat (Y) může pomáhat stát se dobrým sportovcem (A) a zároveň ubírat čas na studium (B).



Obr. 28 Kauzální salát: snaha zafixovat pohlaví, známky i univerzitu najednou. Když kontrolujeme příliš mnoho proměnných, nevyhnutelně mezi nimi bude i collider.

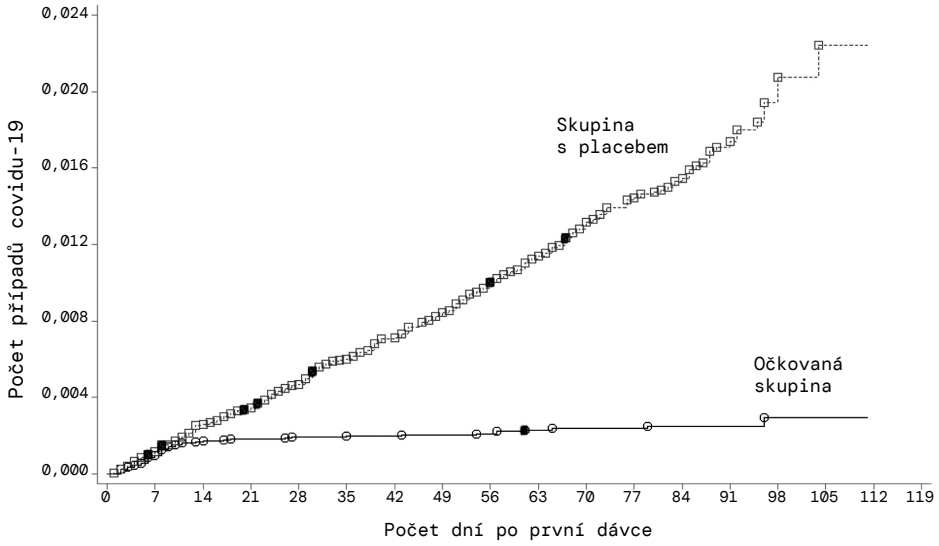
Pravidlo čtvrté

Provádějte experimenty s kontrolní skupinou



Jakýkoliv rozdíl mezi B a B′ musí být výsledkem intervence

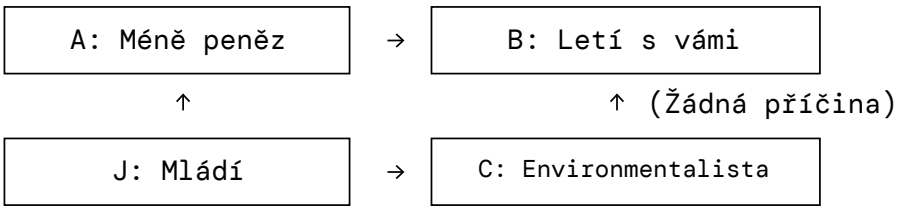
Obr. 29 Kontrolovaný experiment. Obě skupiny jsou stejně vystavené dalším vlivům a liší se jen tím, zda dostanou léčbu A. Jakýkoli rozdíl mezi výsledky B a B´ proto musí být výsledkem intervence.



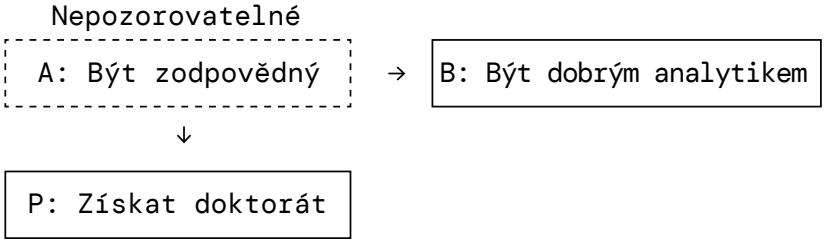
Obr. 30 Případy covidu-19 ve studii vakcíny Pfizer/BioNTech podle počtu dní od první dávky, ve skupině s placebem a v očkované skupině. Zhruba od druhého týdne se křivky rozcházejí. Zdroj: FDA.

Pravidlo čtvrté

Důsledek: nepřehlížejte korelace



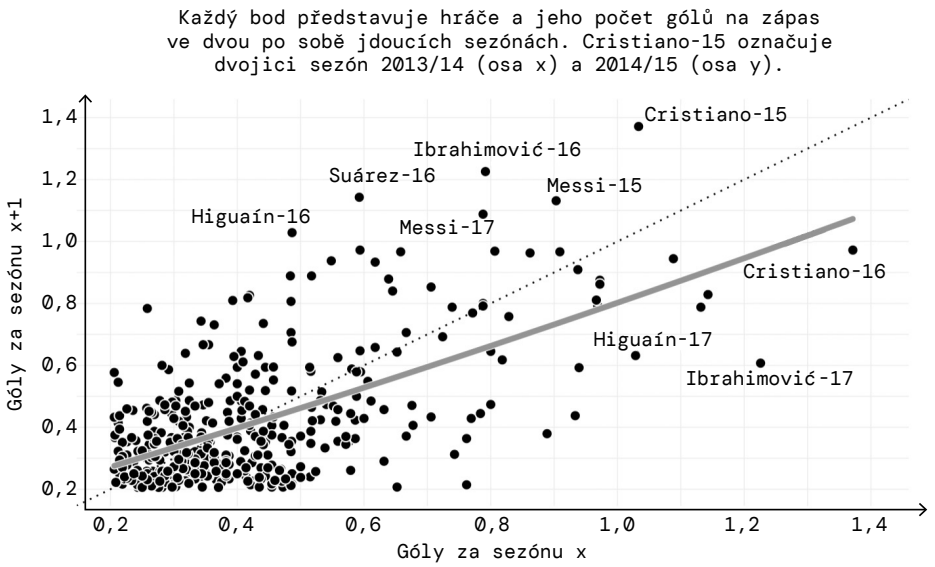
Obr. 31 Mládí (J) znamená méně peněz (A), a proto mladí lidé létají s nízkonákladovou společností (B). Zároveň jsou častěji environmentalisté (C). Mezi B a C tak vzniká souvislost, i když jedno nezpůsobuje druhé.



Obr. 32 Zodpovědnost (A) nelze přímo pozorovat. Dělá z lidí dobré analytiky (B) a zároveň jim pomáhá získat doktorát (P). Doktorát proto dobrého analytika předpovídá, i když sám příčinou není.

Pravidlo páté

Události tíhnou zpátky k průměru



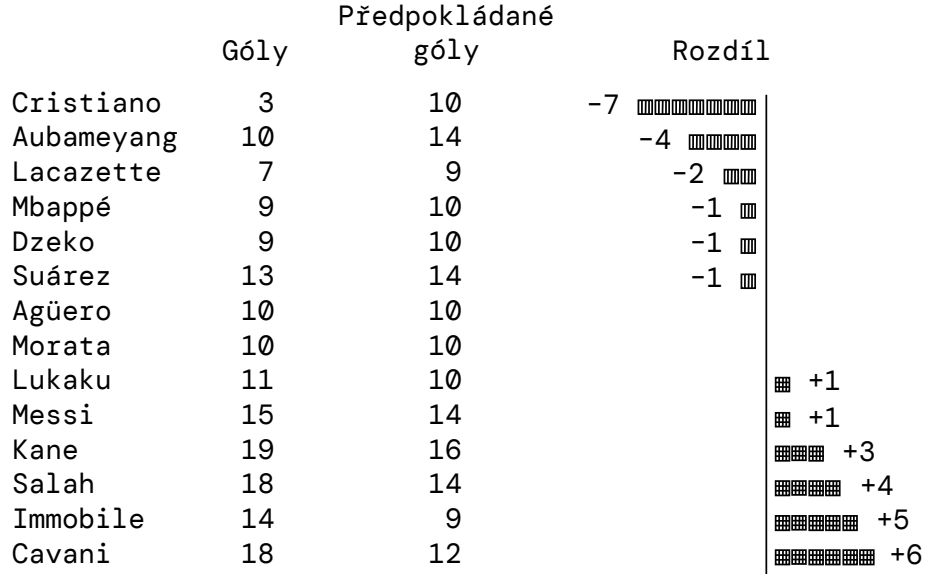
Obr. 33 Góly na zápas u 392 hráčů z pěti nejlepších evropských lig v letech 2014 až 2018. Hráči, kteří v jedné sezóně dávali hodně gólů, dávali v té další obvykle méně.

	2015-2016	2016-2017
Ibrahimović	1,2 (PSG)	0,6 (MAN)
Suárez	1,1 (FCB)	0,8 (FCB)
Higuaín	1,0 (NAP)	0,6 (JUV)

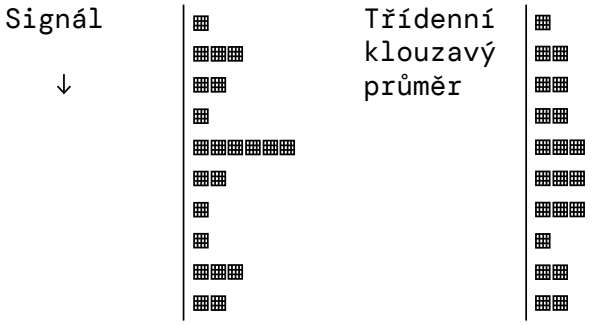
Obr. 34 Góly na zápas tří nejlepších střelců v sezónách 2015/16 a 2016/17. Ibrahimović, Suárez ani Higuaín svou úspěšnost v další sezóně nezopakovali.

Pravidlo páté

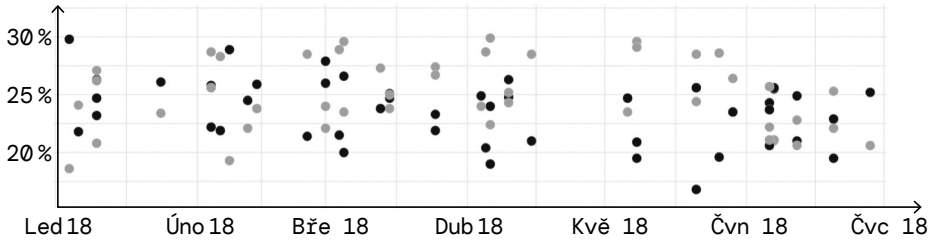
Nepleťte si signál a šum



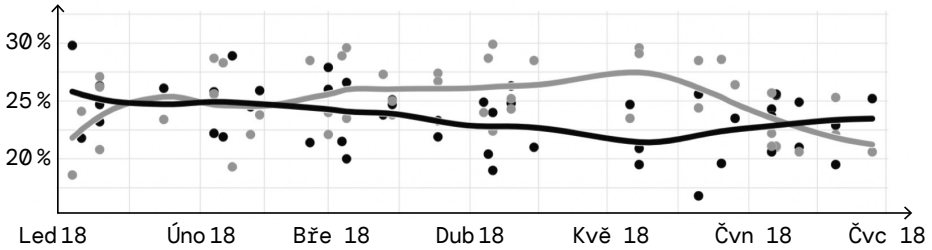
Obr. 35 Vstřelené a předpokládané góly nejlepších střelců v lize k lednu 2018. Cristiano Ronaldo dal o 7 gólů méně, než odpovídalo jeho šancím. Zdroj: Opta.



Obr. 36 Váha měřená každý den (signál) a její třídenní klouzavý průměr. Průměr tlumí náhodné výkyvy, a proto je z něj skutečný vývoj lépe vidět.



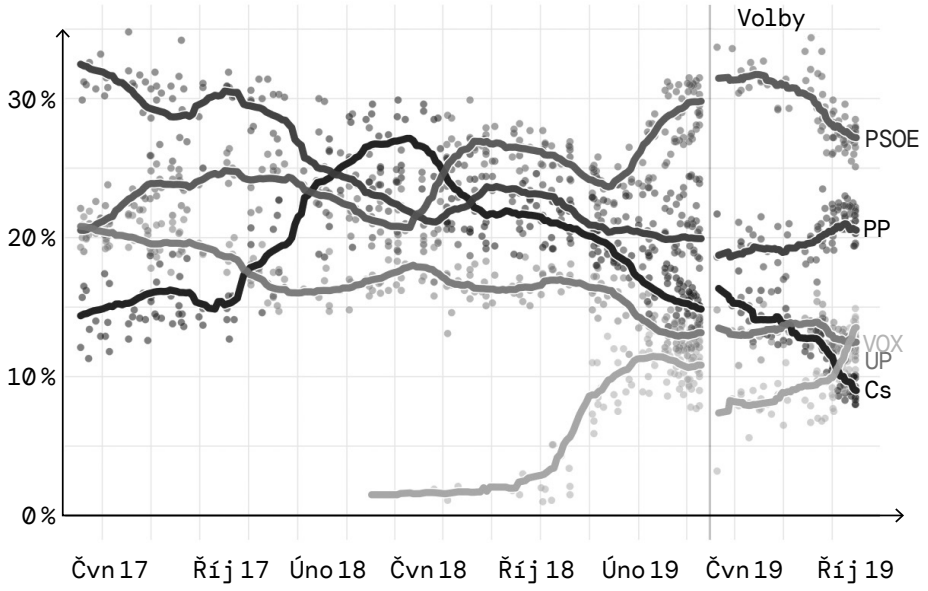
Obr. 37 Odhady volebních preferencí Lidové strany (PP, černé body) a Ciudadanos (šedé body) na začátku roku 2018 podle desítek průzkumů. Z jednotlivých průzkumů skutečný signál rozpoznat nešel.



Obr. 38 Tytéž průzkumy zprůměrované. Ciudadanos posilovaly až do konce května 2018. Po úspěšném hlasování o nedůvěře premiéru Rajoyovi začaly ztrácet ve prospěch PP.



Obr. 39 Dilema filtrování. Třídenní klouzavý průměr odstraní šum, ale skutečný signál zachytí se zpožděním.



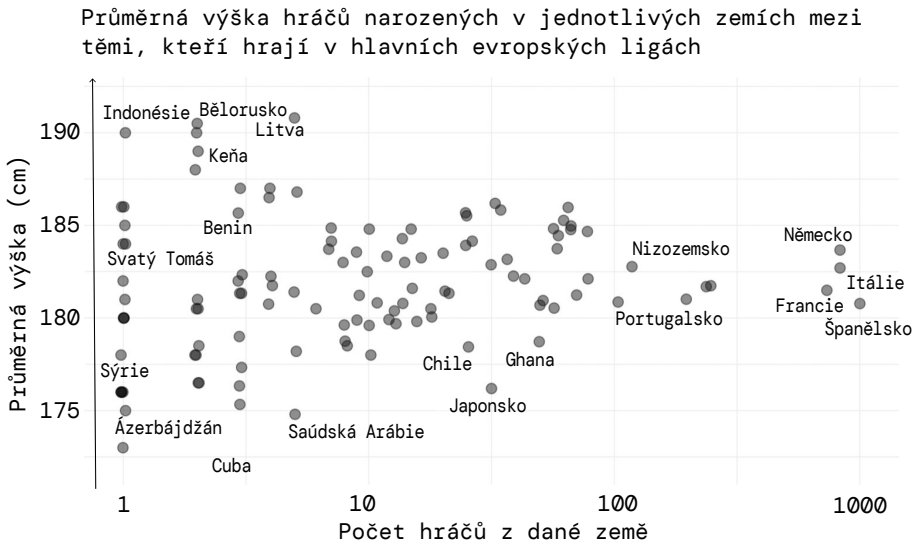
Obr. 40 Průměrný vývoj volebních průzkumů ve Španělsku v letech 2017 až 2019. Každý bod představuje jeden průzkum, čáry znázorňují průměry. Volby se konaly v dubnu a v listopadu 2019.

Typická chyba průzkumu	±3,5 bodu ▦ ▦ ▦ ▦ ▦ ▦ ▦
Typická průměrná chyba průzkumu	±2 body ▦ ▦ ▦ ▦

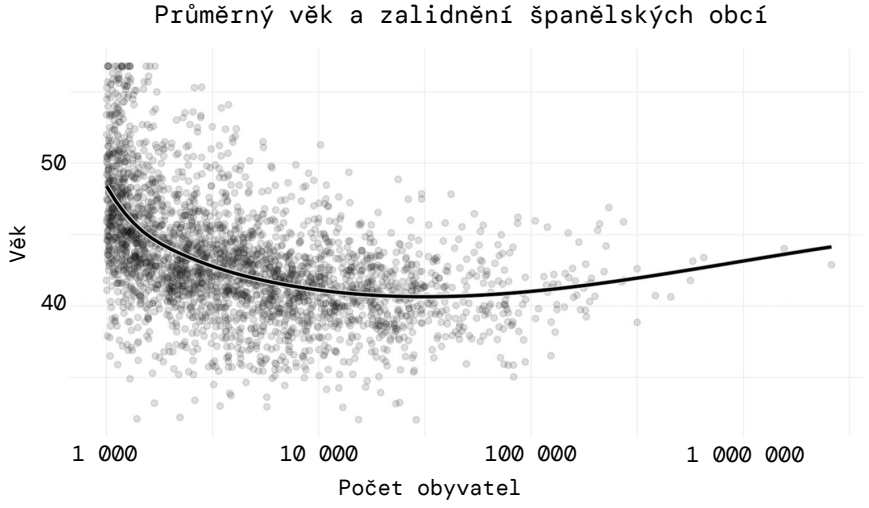
Obr. 41 Průměrná chyba volebních průzkumů ve Spojených státech podle analýzy 4 000 průzkumů, kterou vedl statistik Andrew Gelman. Jeden průzkum se mýlí typicky o 3,5 bodu, průměr více průzkumů zhruba o 2 body.

Pravidlo páté

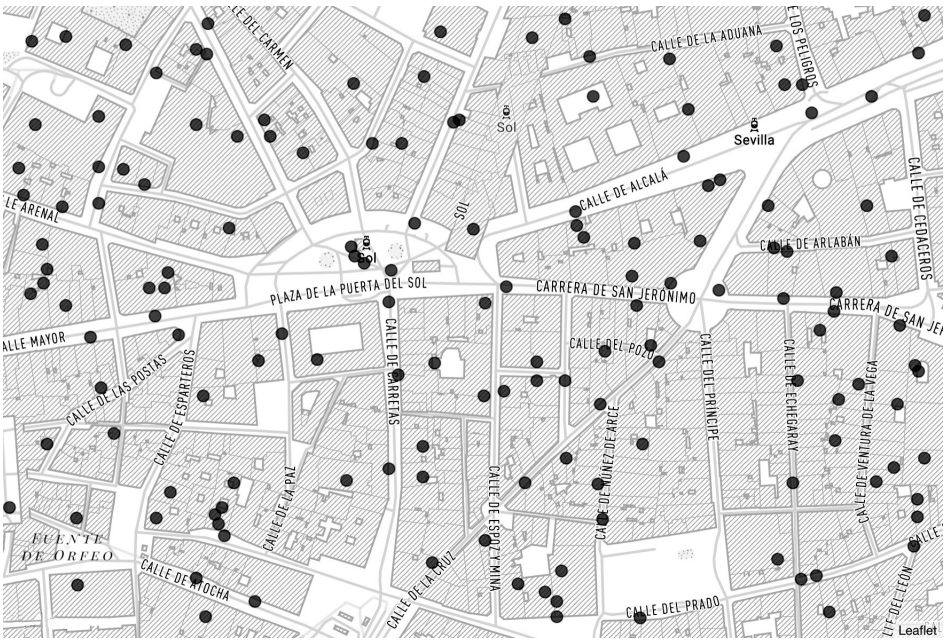
Máte-li málo dat, kroťte své nadšení



Obr. 42 Průměrná výška hráčů z jednotlivých zemí podle počtu hráčů z dané země. Graf má trychtýřovitý tvar: u zemí s malým počtem hráčů je větší rozptyl. Indonésii zastupuje jediný hráč, brankář Emil Audero.



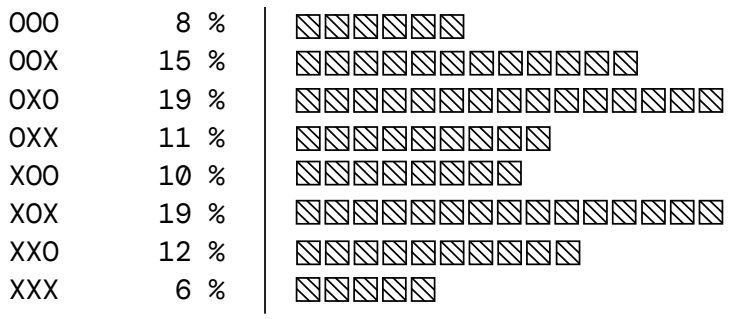
Obr. 43 Průměrný věk a počet obyvatel španělských obcí. Malé vesnice jsou nejstarší a věk se v nich liší mnohem víc než ve městech.



Obr. 44 Mapa centra Madridu se 150 body. Zkuste v nich najít nějaký vzorec. Jak autor v kapitole vysvětluje, body rozmístil po mapě náhodně.

```
OOO    12,5 %  ▧▧▧▧▧▧▧▧▧▧
OOX    12,5 %  ▧▧▧▧▧▧▧▧▧▧
OXO    12,5 %  ▧▧▧▧▧▧▧▧▧▧
OXX    12,5 %  ▧▧▧▧▧▧▧▧▧▧
XOO    12,5 %  ▧▧▧▧▧▧▧▧▧▧
XOX    12,5 %  ▧▧▧▧▧▧▧▧▧▧
XXO    12,5 %  ▧▧▧▧▧▧▧▧▧▧
XXX    12,5 %  ▧▧▧▧▧▧▧▧▧▧
```

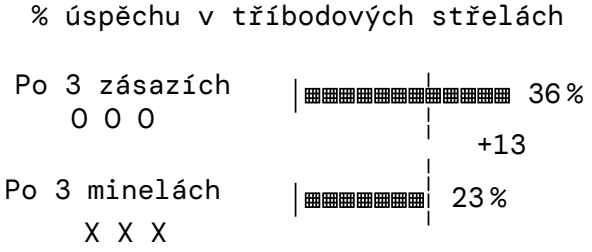
Obr. 45 Četnost trojic v hodech mincí generovaných počítačem (O hlava, X orel). Všechny trojice se vyskytují stejně často.



Obr. 46 Četnost trojic v sekvencích, které napsalo 91 lidí. Lidé psali zhruba o polovinu méně trojic OOO a XXX, než odpovídá náhodě, a nadužívali trojice OXO a XOX.

Pravidlo páté

Pozor na série nezdarů



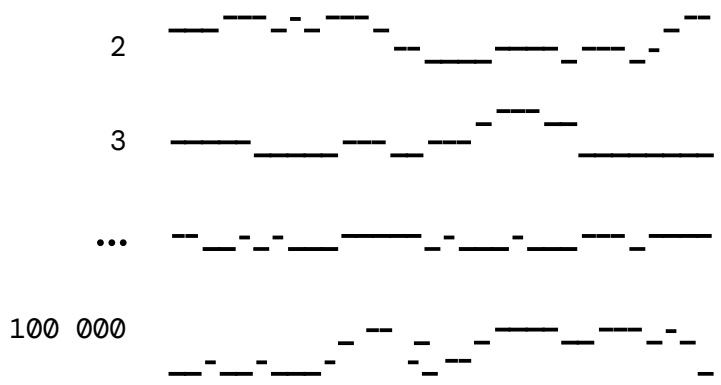
Obr. 47 Úspěšnost tříbodových střel po třech proměněných a po třech neproměněných pokusech: 36 % oproti 23 %.

Pravidlo páté

Rozlišujte mezi průzkumem a potvrzením



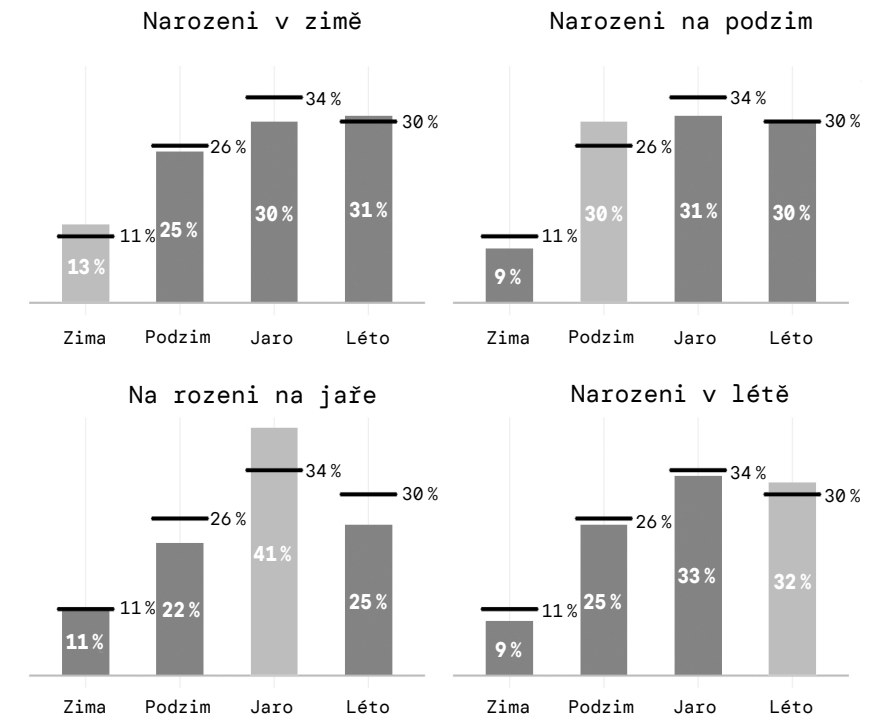
Obr. 48 Řada 32 datových bodů, například počet návštěv posilovny v jednotlivých týdnech roku 2022.



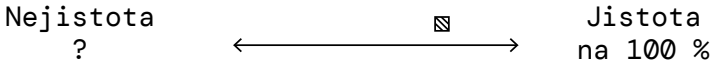
Obr. 49 Dalších 100 000 datových řad, od průměrné teploty v Benidormu po počet japonských turistů v Barceloně. Mezi tolika řadami se vždy najde nějaká, která s tou první náhodou koreluje.

	S hypotézou	Bez hypotézy
Podívat se data jednou	potvrdit	zkoumat
Podívat se data n-krát	zkoumat	zkoumat

Obr. 50 Kdy data potvrzují a kdy je jen zkoumáte. Potvrdit hypotézu lze jen tehdy, když ji máte předem a podíváte se na data jednou.

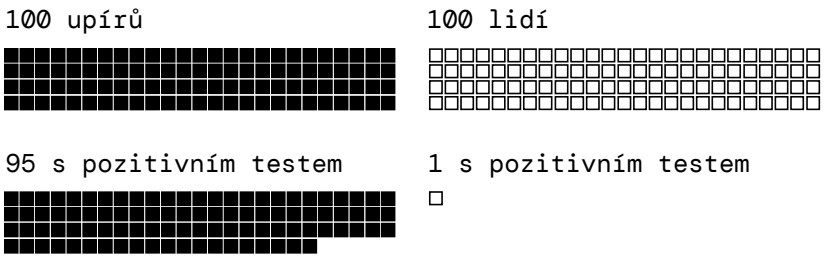


Obr. 51 Oblíbené roční období lidí narozených v jednotlivých ročních obdobích (sloupce) a oblíbené roční období celkově (čáry). Jaro uvedlo 34 % všech respondentů, ale 41 % těch, kteří se na jaře narodili.

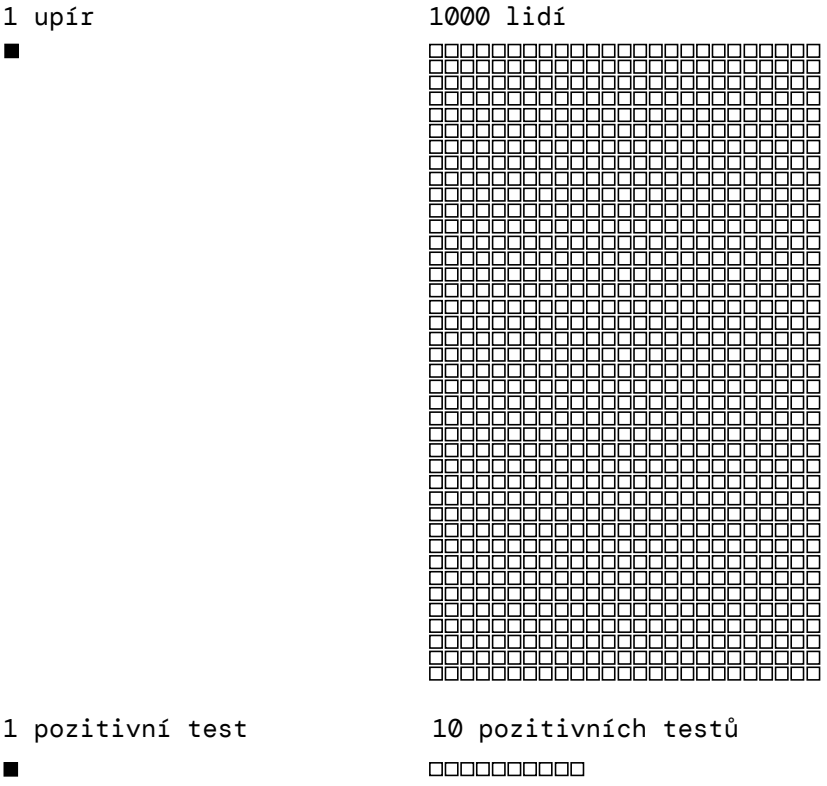


Obr. 52 Potvrzování hypotézy není skok z nejistoty do jistoty. Je to kontinuum, po kterém se postupně posouváme k vyšší pravděpodobnosti.

Přijměte, že nejistota vládne



Obr. 53 Test na upíry v populaci 100 upírů a 100 lidí. Pozitivní test má 95 upírů a 1 člověk, takže 99 % pozitivních testů patří skutečným upírům (95 z 96).



Obr. 54 Tentýž test, když připadá jeden upír na 1 000 lidí. Pozitivní test má 1 upír a 10 lidí, takže skutečným upírům patří jen 8,7 % pozitivních testů (1 z 11).

$$P(\blacksquare|\text{posit.}) = \frac{P(\blacksquare) \times P(\text{posit.}|\blacksquare)}{P(\blacksquare) \times P(\text{posit.}|\blacksquare) + P(\square) \times P(\text{posit.}|\square)}$$

Obr. 55 Bayesova věta pro test na upíry (■ upír, □ člověk). Pravděpodobnost, že je člověk s pozitivním testem upír, závisí na spolehlivosti testu i na tom, kolik upírů v populaci je.

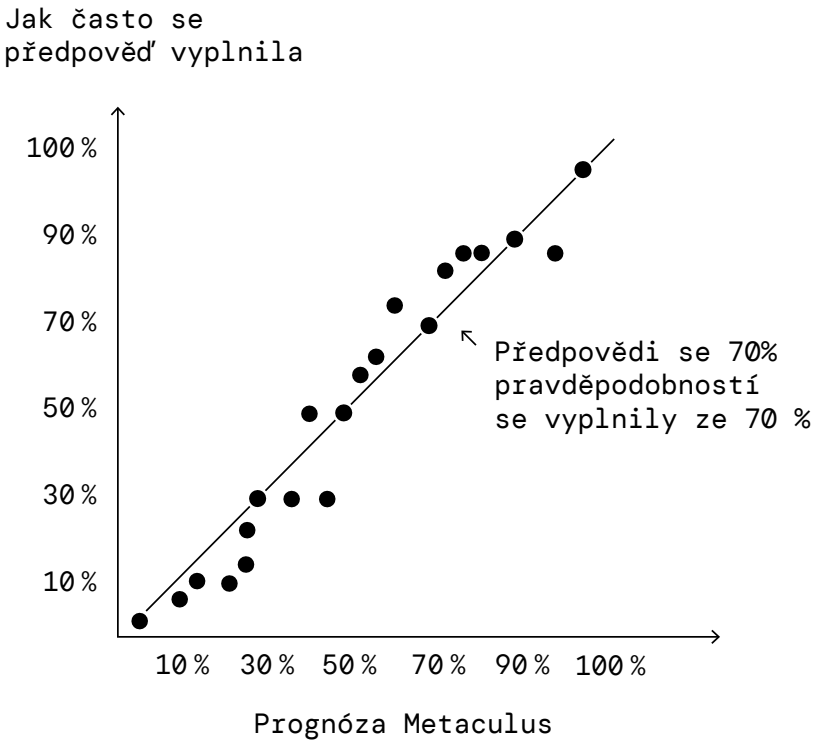
Myslete jako superprognostik

Pravděpodobnost	
< 5 %	Téměř nemožné
10 %	Velmi nepravděpodobné
20 %	
25 %	Nepravděpodobné
35 %	
40 %	Reálná možnost
50 %	
55 %	Pravděpodobné
65 %	
75 %	
80 %	Velmi pravděpodobné
85 %	
90 %	
> 95 %	Téměř jisté

Obr. 56 Jak britští epidemiologové během pandemie převáděli slovní odhady na konkrétní pravděpodobnosti.

Prognostik A:							
Předpověď:	99 %	99 %	99 %	99 %	99 %	99 %	99 %
Úspěch:	x	o	o	o	o	x	o
Prognostik B:							
Předpověď:	60 %	60 %	60 %	60 %	60 %	60 %	60 %
Úspěch:	x	o	o	o	x	o	o

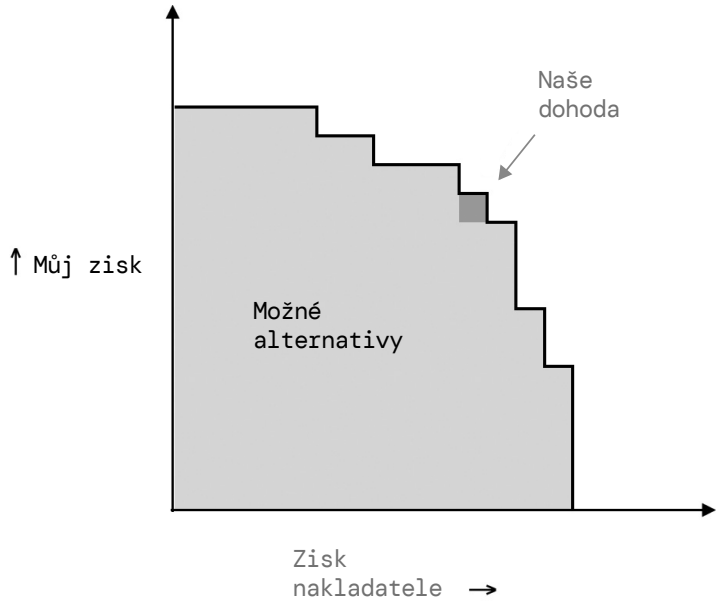
Obr. 57 Dva hypotetičtí prognostici se stejnou přesností 71 %. Prognostik A je špatně kalibrovaný: co se podle něj mělo stát jen v 1 % případů, se stalo v 29 %.



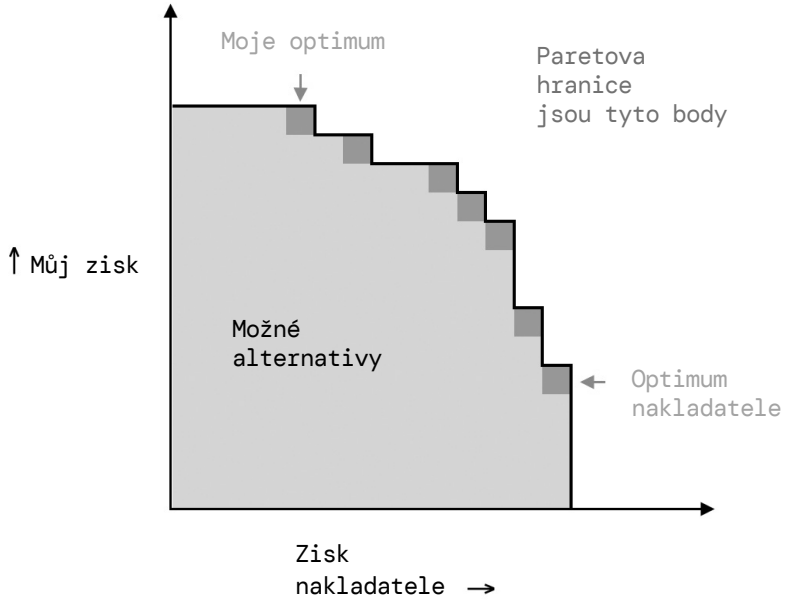
Obr. 58 Kalibrace předpovědí na platformě Metaculus. Předpovědi se 70% pravděpodobností se vyplnily ze 70 %, body proto leží přibližně na diagonále. Zdroj: Metaculus, 10. dubna 2022.

Pravidlo sedmé

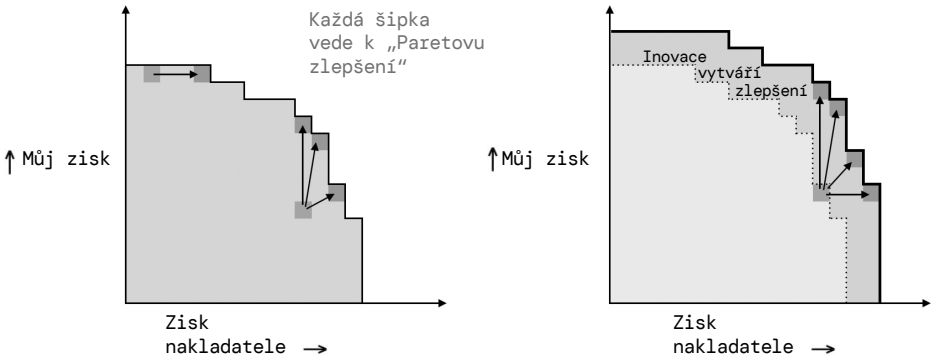
Používejte ve svém uvažování Paretův princip



Obr. 59 Dělení honoráře mezi autora a nakladatele. Světle šedě jsou vyznačené možné alternativy, tmavě šedě dohoda, které dosáhli.



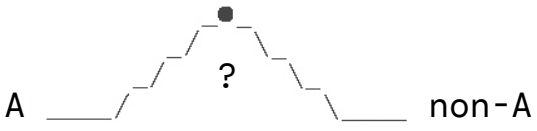
Obr. 60 Paretova hranice. Každý bod na této křivce představuje situaci, kdy si žádná ze stran nemůže polepšit, aniž by se druhé straně vedlo hůř.



Obr. 61 Paretovo zlepšení. Každá šipka vede k řešení, při kterém si aspoň jedna strana polepší. Inovace, například český překlad knihy, posouvá hranici toho, co je možné.

Pravidlo osmé

Zkroťte ukvapené závěry



Obr. 62 Úsudek je jako míč na kopci. Stačí málo informací a míč se skutálí k jednomu ze závěrů: A, nebo non-A.

G L _ _
_ _ T E R _ _ R E
P _ _ N

Obr. 63 Neúplná anglická slova z experimentu týmu Emily Proninové. Zkuste rychle a bez dlouhého přemýšlení doplnit chybějící písmena.

Doporučená literatura

Na začátku knihy píšu o fascinujícím životě úhořů, inspirován skvělým dílem Patrika Svenssona *The Book of Eels*.

V prvních kapitolách používám několik příkladů z knihy *The Genetic Lottery: Why DNA Matters for Social Equality* od Kathryn Paige Hardenové, Princeton University Press, 2022. Je to kniha o vztahu mezi geny a prostředím. Považuji ji za zajímavou, protože se zaměřuje na originální úhel pohledu: stejně jako může být nespravedlivé být chudý nebo bohatý, může být nespravedlivé i narodit se s určitými genetickými variantami, které vám dávají výhodu, například být atraktivní nebo opatrný.

Nejlepší úvod do práce Daniela Kahnemana a Amose Tverského najdete v knize Michaela Lewise *The Undoing Project*, Penguin, 2017. Jejich myšlenky jsou dále rozpracovány v knize Daniela Kahnemana *Myšlení, rychlé a pomalé*, Penguin, 2011.

Pokud se chcete dozvědět více o kauzalitě, doporučuji knihu *Causal Inference: What If* od Miguela Hernána a Jamese Robinse, Taylor and Francis, 2011. Také *The Book of Why* od Judea Pearla a Dany Mackenziové, Penguin, 2019.

V několika kapitolách cituji Richarda McElreatha, autora knihy *Statistical Rethinking*, Chapman & Hall/CRC, 2020. Je to matematická učebnice bayesovské statistiky, ale je vynikající. Pokud umíte číst vzorce, je to fascinující četba, i když vás tato oblast statistiky primárně neláká.

Příběh Leonida Kantoroviče jsem objevil v knize Francise Spufforda *Red Plenty*, Faber and Faber, 2010. Je to částečně román, částečně esej, plný zajímavých scén; skvělá a jedinečná kniha o historickém konceptu a období kolem roku 1950, kdy plánovaná ekonomika slibovala zázrak hojnosti.

Chcete-li se dozvědět více o práci Philipa E. Tetlocka a Barbary Mellersové v oblasti předpovídání, můžete si přečíst knihu Tetlocka a Dana Gardnera *Superforecasting: The Art and Science of Prediction*, Random House, 2016.

Úplný seznam zdrojů najdete na webu autora: kiko.llaneras.es