

Dataskеpticism

Mám data.

Data samotná jsou v pořádku.

**Data interpretuji do formálně
korektní vizualizace.**

Něco je špatně, lidé se smějí.

Proč?



Otázky

Mám opravdu kompletní data?

Nepřehlížím něco?

**Nezapomněl jsem na nějaký
skrytý vliv?**

**Neudělal jsem nějakou naivní
chybu?**

**Nemám ve skutečnosti tak
trochu dada?**

korelace

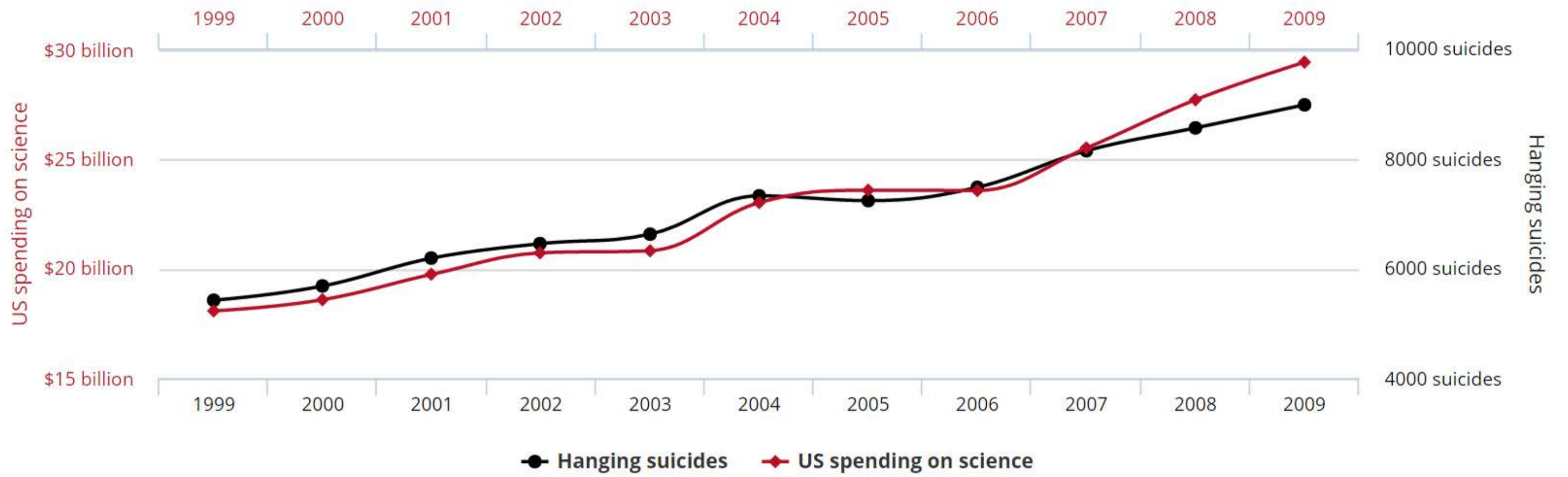


kauzalita

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



Correlation: 99.79% (r=0.99789126)



tylervigen.com

Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

→ **Spurious Correlations**

ceteris paribus

**„Ostatní stejné“ — zafixujeme
všechny veličiny kromě těch,
které sledujeme.**

Vynikající v laboratorních podmínkách.

V reálném světě princip ceteris paribus prakticky nefunguje.

známé známé

data

známé neznámé
hypotéza

neznámé známé

lidé neznají

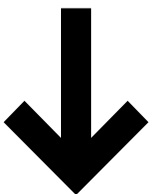
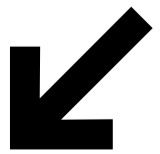
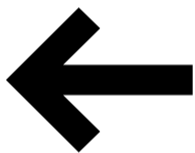
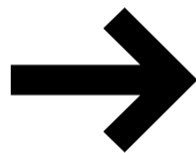
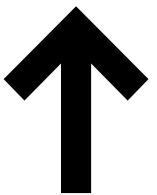
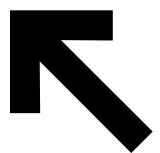
sami sebe

neznámé neznámé

!!!

V datech nemusí být všechno.

Vždy si uvědomte, že existují věci, jevy, skutečnosti, o kterých nevíte, že je nevíte. Mohou být (a často jsou) ve skutečnosti důležitější, než to co víte.

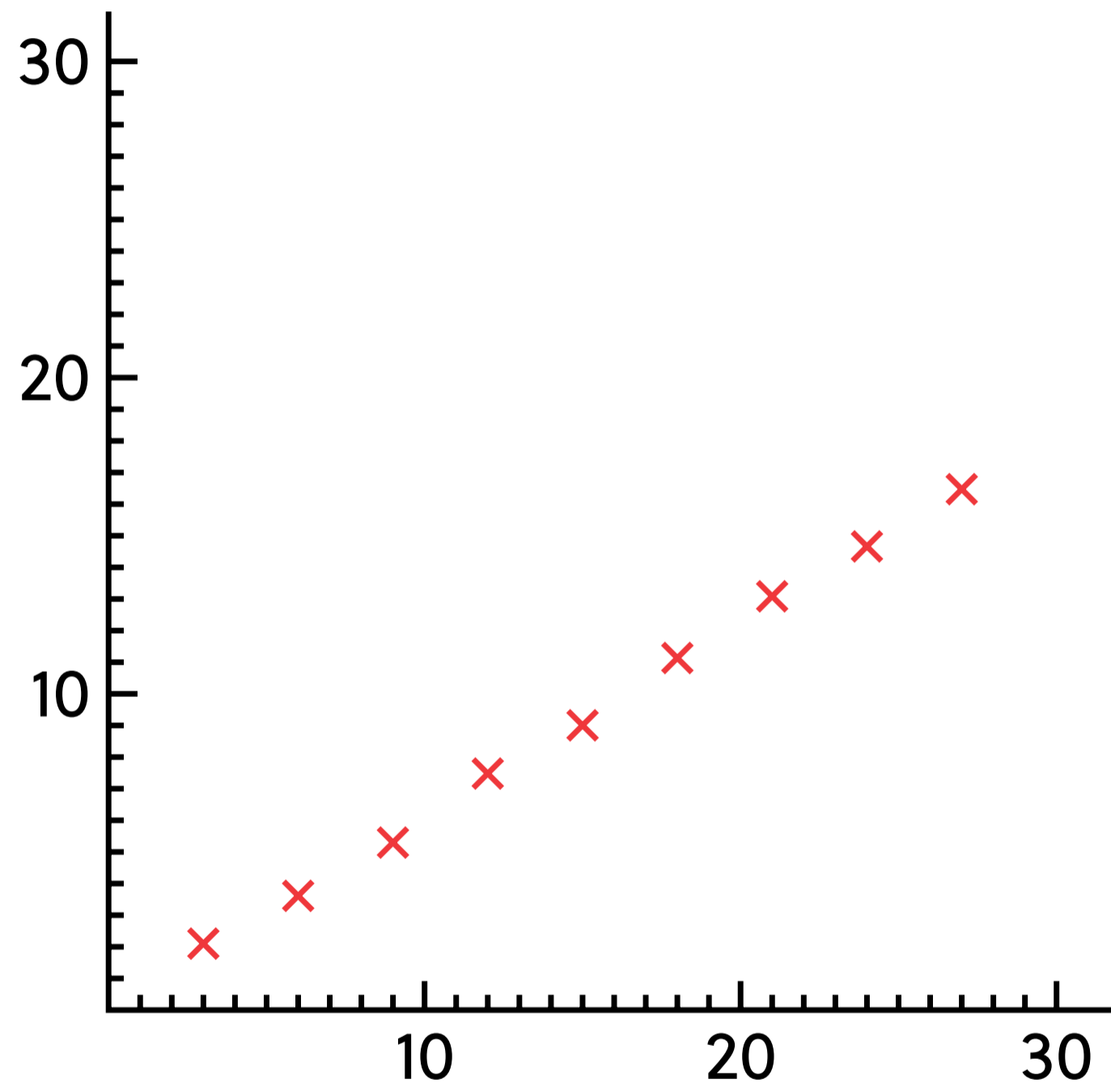


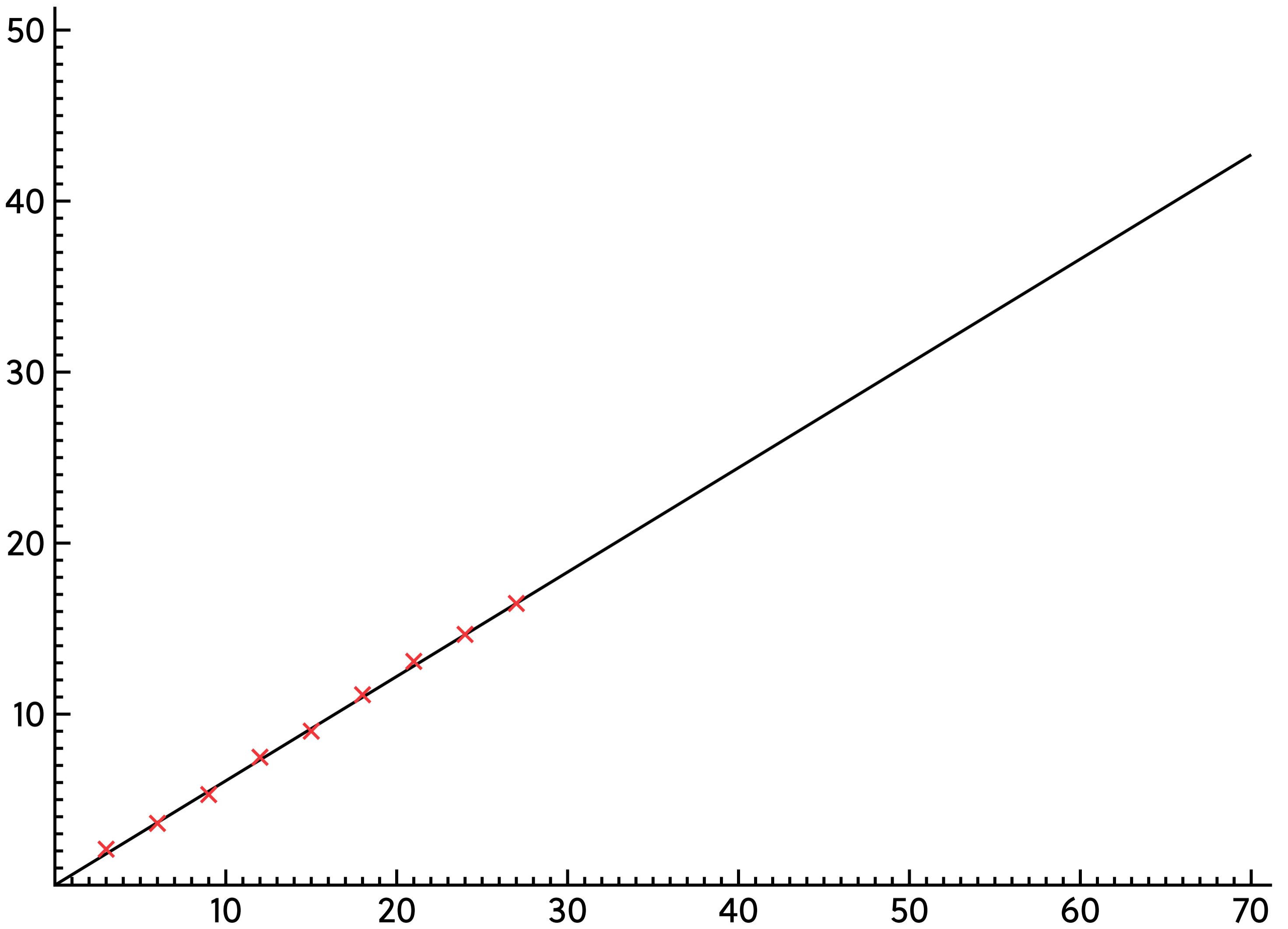
**Neznámé neznámé může
být subjektivní.**

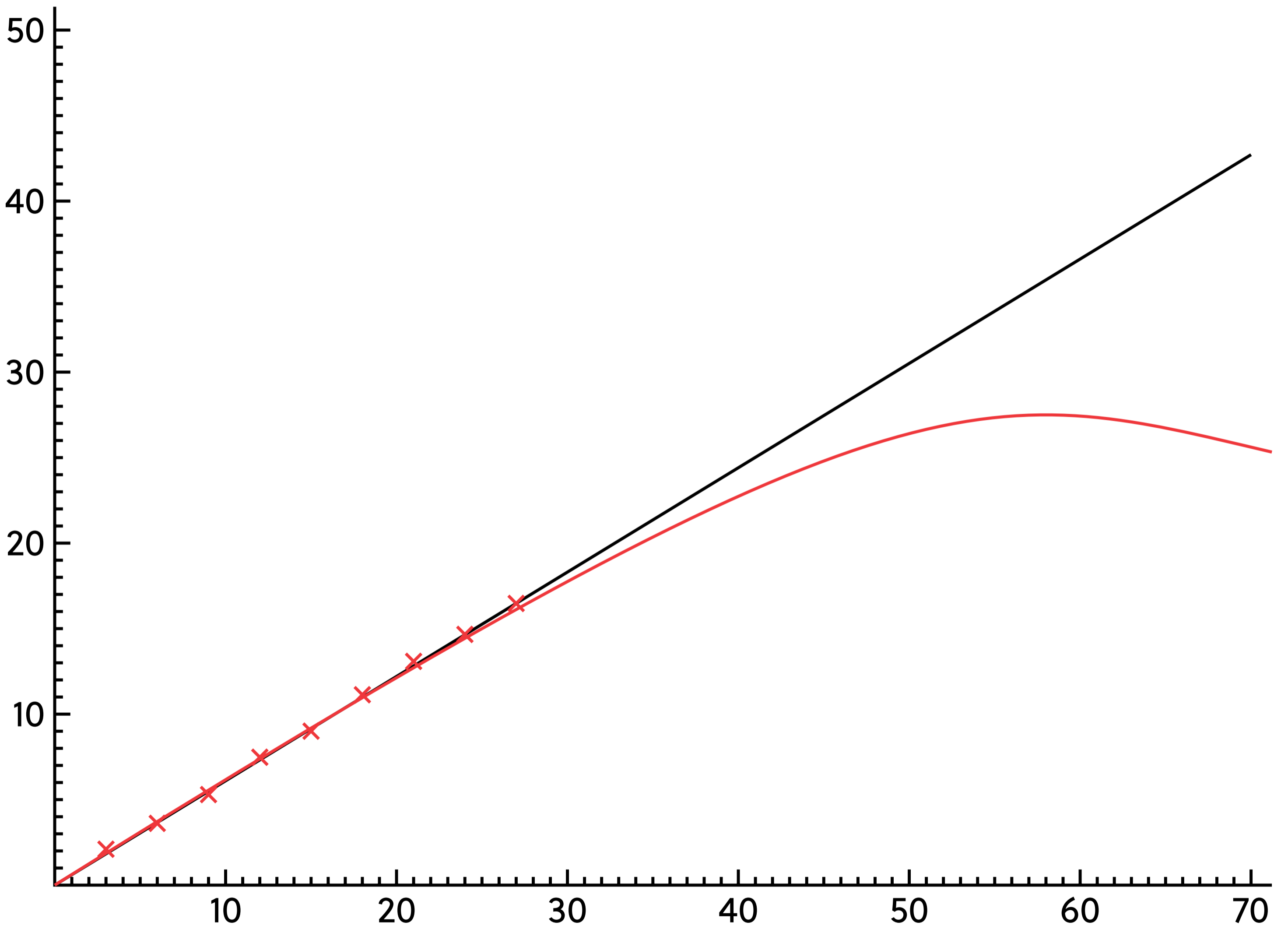
**To, že nevíme, že něco nevíme,
nutně neznamená, že to neví
někdo jiný.**

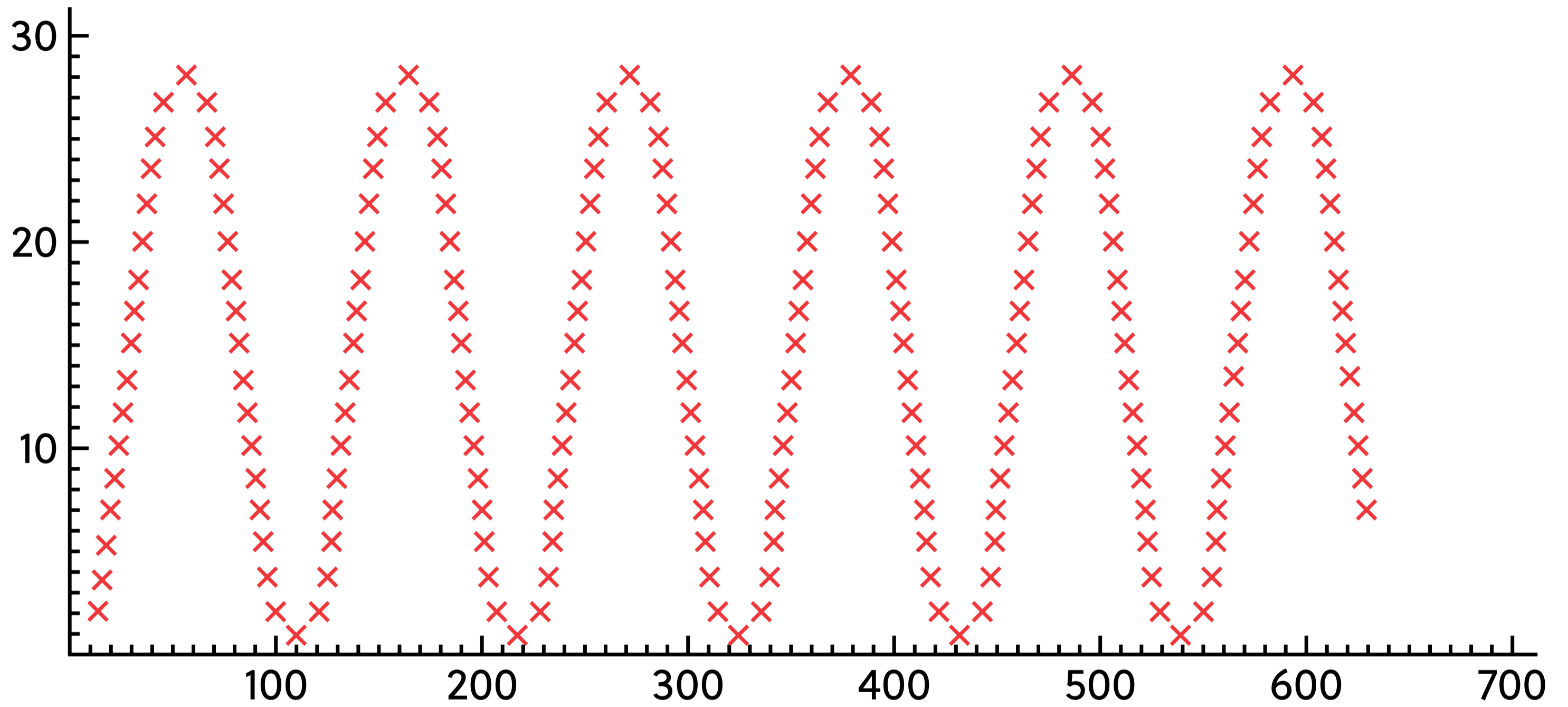
Implikace:

**Nesnažte se predikovat/
neextrapolujte na základě
historických dat.**





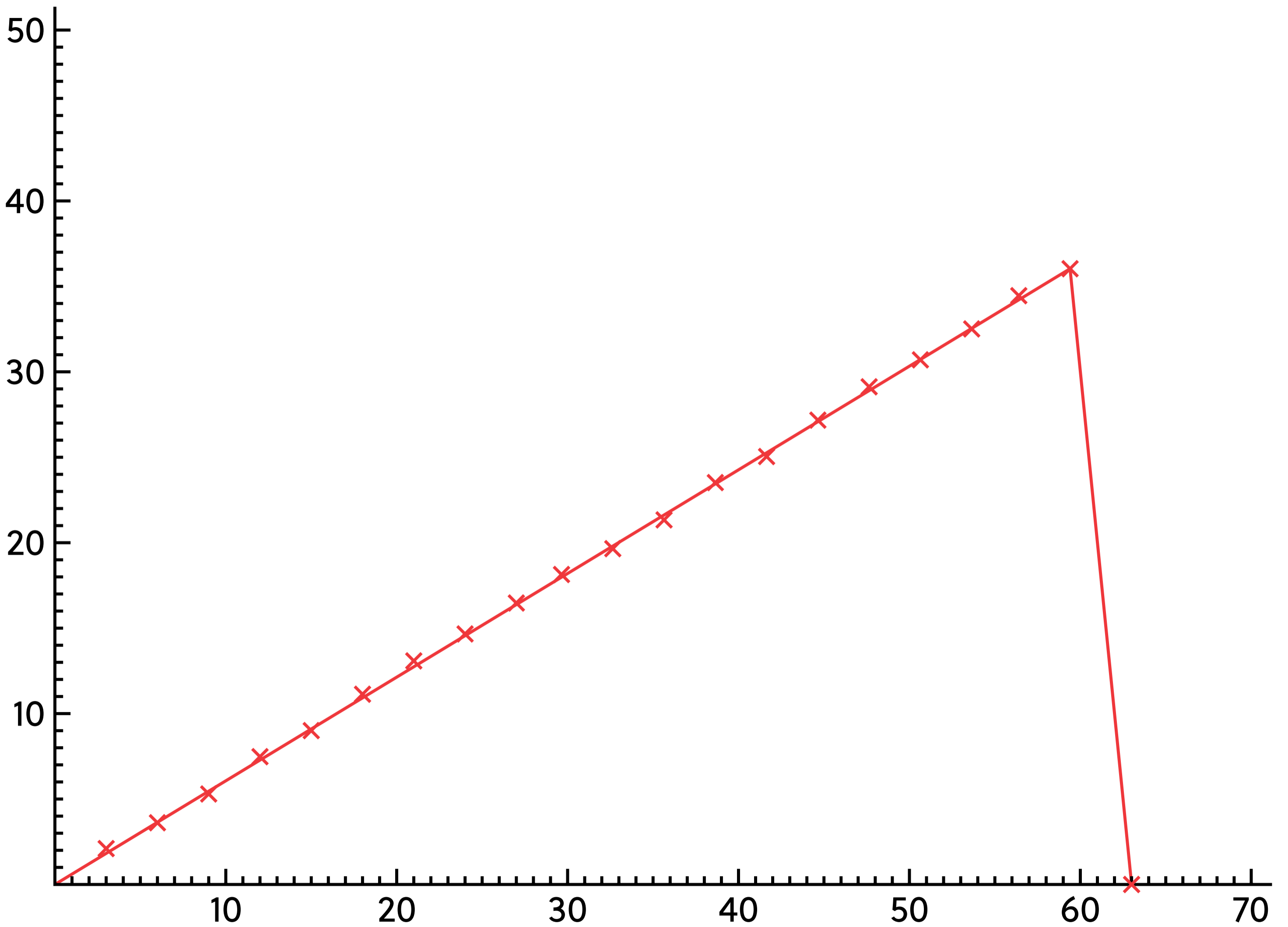




Nejužitečnější problém

Abychom poznali jaké rozložení data mají, musíme mít dostatek datapointů.

Abychom poznali, kolik je dostatek datapointů, musíme znát rozložení dat.



→ **Confirmation bias**

8

Pokladník charitativní společnosti si v červenci udělal tabulku, aby viděl, kolik peněz přibývá na konto každý měsíc. Řekl si: „Věřím, že to půjde tak dál,“ a dopsal do tabulky další tři čísla. Jaká?

měsíc	květen	červen	červenec	srpen	září	říjen
stav konta [v milionech Kč]	1,5	1,7	1,9			



→ **America's epidemic of unnecessary care**

zpětná

interpretace

klam narativity

synchronicita

The narrative fallacy addresses our limited ability to look at sequences of facts without weaving an explanation into them, or, equivalently, forcing a logical link, an arrow of relationship upon them. Explanations bind facts together. They make them all the more easily remembered; they help them make more sense. Where this propensity can go wrong is when it increases our impression of understanding.

— Nassim Taleb, Černá labuť

**problém
přeživších**

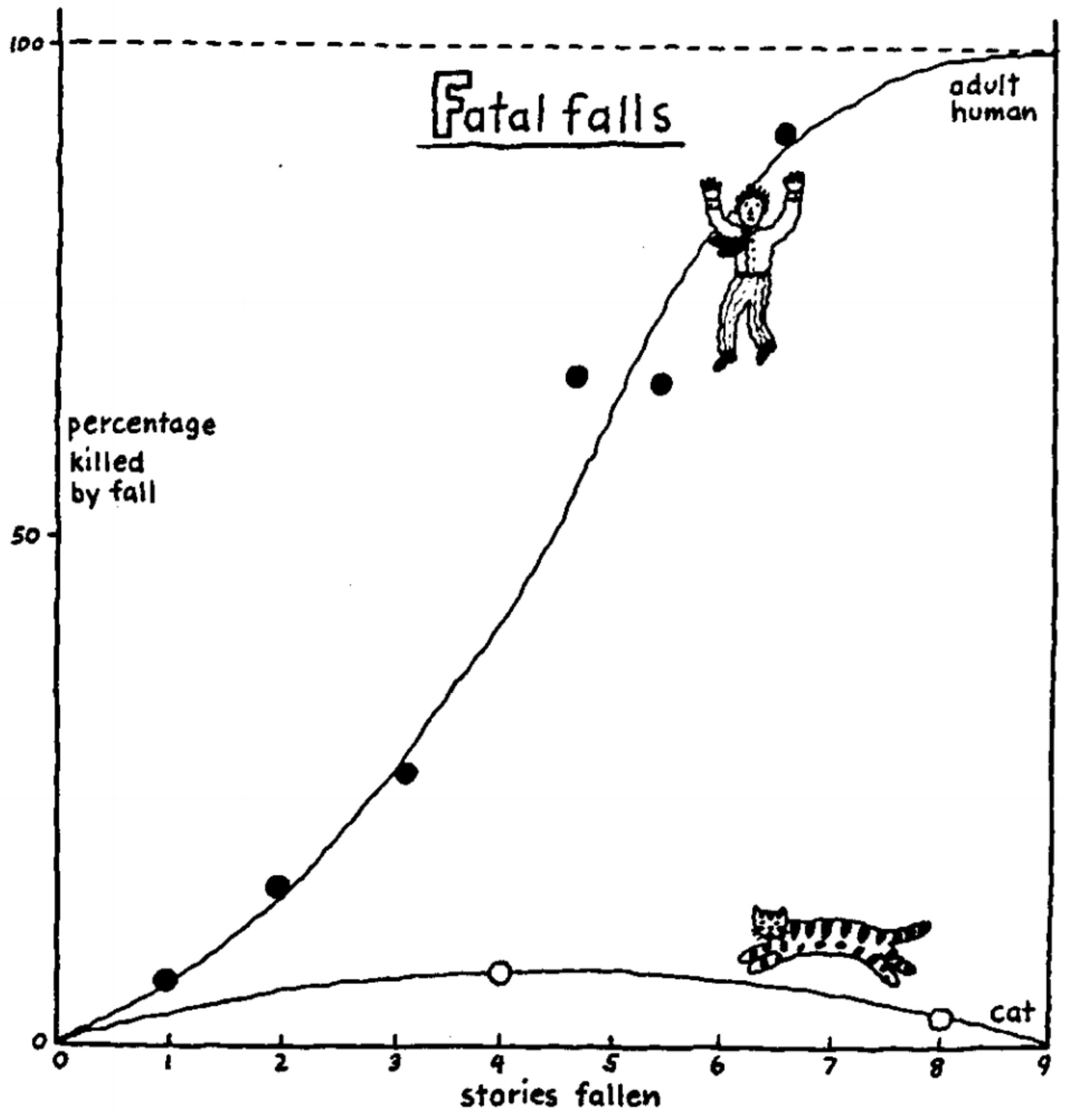
Diagoras, who is called the atheist, being at Samothrace, one of his friends showed him several pictures of people who had endured very dangerous storms; "See," says he, "you who deny a providence, how many have been saved by their prayers to the Gods."

"Ay," says Diagoras, "I see those who were saved, but where are those painted who were shipwrecked?"

— Cicero, On the Nature of the Gods



In a study performed in 1987 it was reported that cats who fall from less than six storeys, and are still alive, have greater injuries than cats who fall from higher than six storeys. It has been proposed that this might happen because cats reach terminal velocity after righting themselves at about five storeys, and after this point they relax, leading to less severe injuries in cats who have fallen from six or more storeys.



Another possible explanation for this phenomenon would be survivorship bias. Cats that die in falls are less likely to be brought to a veterinarian than injured cats, and thus many of the cats killed in falls from higher buildings are not reported in studies of the subject.

— [Wikipedia: Survivorship bias](#)

big Dada

FEB 16, 2012 @ 11:02 AM 2,850,601 VIEWS

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



Kashmir Hill, FORBES STAFF

Welcome to The Not-So Private Parts where technology & privacy collide

[FOLLOW ON FORBES \(2081\)](#)



Opinions expressed by Forbes Contributors are their own.

FULL BIO ▾



TARGET

Target has got you in its aim

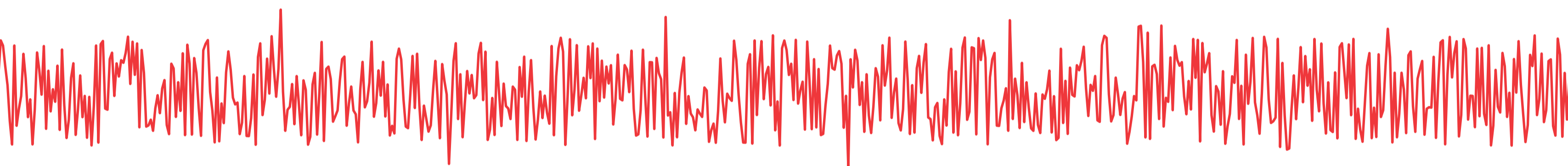
Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying

→ **Big data: are we making a big mistake?**

šum

×

signál



Co si odnést

Nezapomeňte na to, že existuje svět i mimo vaše data.

Náhoda může hrát mnohem větší roli než si myslíte.

Kam dál?

Nassim Nicholas Taleb

— **Zrádná nahodilost**

— **Černá labuť**

— **Antifragilita**