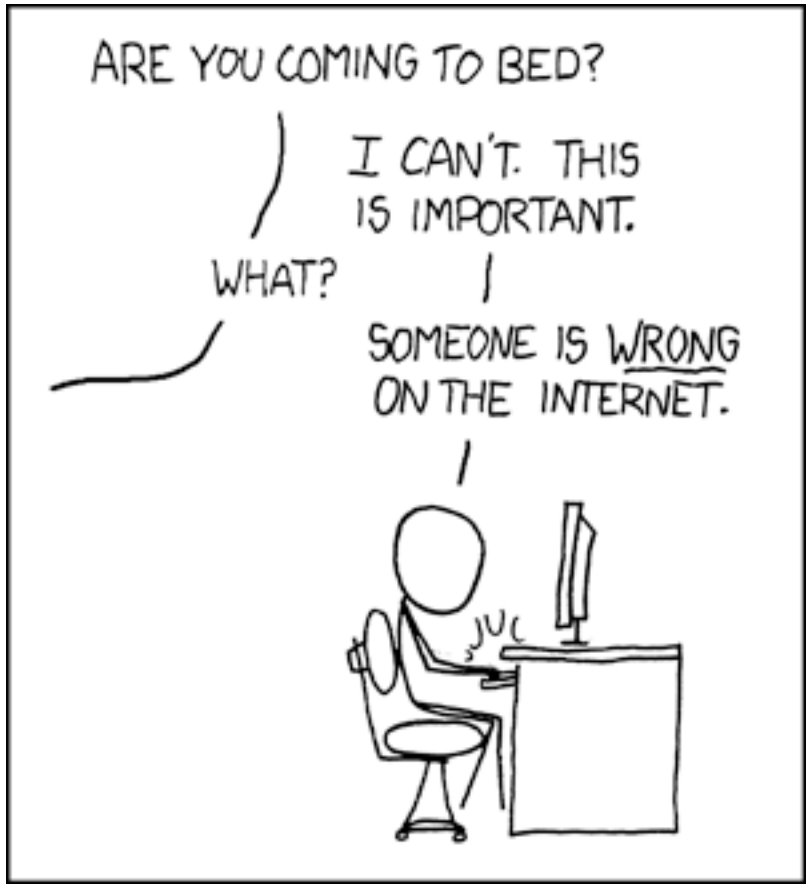


How to Lie with Statistics

312 23Su
Lecture 23

https://xkcd.com/386/

Acknowledgements

"How to Lie with Statistics" by Darrell Huff (1954)

Previous versions of this lecture by Stefano Tessaro, Anna Karlin and Alex Tsun

"Calling Bullshit: Data Reasoning in a Digital World" by Carl Bergstrom and Jevin West

"Naked Statistics: Stripping the Dread from Data" by Charles Wheelan

And more!

Statistical Inference

Making an estimate or prediction about a **population** based on a **sample**

Impossible/impractical to survey an entire population

Random unbiased samples used to draw conclusions, with some margin of error

Sampling

"The Literary Digest" Magazine poll to predict 1936 election.

Alfred Landon vs. Franklin D Roosevelt

10 million surveys, 2.4 million responses

subscribers, owners of cars and telephones on a List

Electoral Votes	Prediction	Actual
Landon	370	8
Roosevelt	161	523

Sampling

Not representative

- Voluntary response bias (24% responded)
- Not the right population (more money/education/info than average American)

Not random

- Convenience sampling

https://xkcd.com/1999/

Sampling

In more recent years…

Who answers phone calls? Democrats (especially during Covid!), certain demographics, more politically engaged and … *people with high social trust*

1. Responses of poll takers => assume something about population

2. Differences between sample and population => weight by race, education, gender, etc.

3. From 2016, in the US => people with low social trust may have begun voting significantly differently

Takeaway

More sampling is **not** a solution for a bad sampling technique.

Numbers based on people are affected by… people. Surprise.

Facebook Likes

Let’s consider all the Facebook posts on the web.

How often would you expect the number ‘1’ to be the first digit of the number of likes of the post?

Facebook Likes

Let's consider all the Facebook posts on the web.

How often would you expect the number '1' to be the first digit of the number of likes of the post?

Maybe you'd expect this to be true for about 1/9 (about 11%) of the posts.

The actual answer: around 30% of posts.

Not Just Likes

This phenomenon has been observed in…

- Facebook likes
- Twitter retweets
- Country populations
- River lengths
- Mountain heights
- Stock prices
- Electrical bills
- Street addresses

EVER SINCE I HEARD THE SIMILE "AS NEGLECTED AS THE NINE BUTTON ON THE MICROWAVE" I'VE FOUND MYSELF ADJUSTING COOK TIMES.

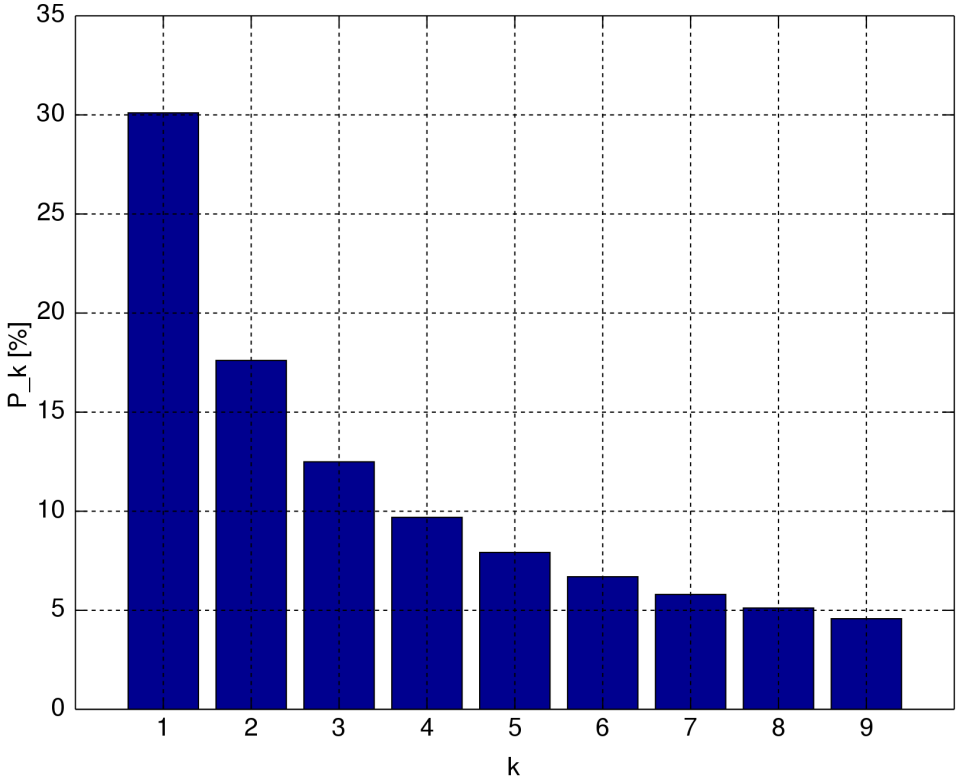
https://xkcd.com/1103/

Benford's Law

In many real-world datasets, the leading digit is likely to be small.

To be precise, a set of numbers is said to fulfill Benford's law if the leading digit $d \in \{1, \dots, 9\}$ occurs with probability

$$\mathbb{P}(d) = \log_{10}(d+1) - \log_{10}(d)$$



By Gknor - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=4509760

Not Just For Fun!

Daniel D. Dorrell

Leading digits of bank accounts

Convicted Wesley Rhodes, a financial advisor – > financial statements

Faked the data.

Admissible in federal, state and local courts of law.

Not Just For Fun!

Bogus retweets, bot networks of followers

Greece manipulating macroeconomic data in application to join the Eurozone

Vote-rigging in Iran's 2009 presidential election

Why?

Organic processes generate numbers that favor small leading digits, whereas naive methods of falsifying data do not.

Indifferent to units of measure.

Why?

Ted Hill: if you pick random numbers from different probability distributions, then they will tend toward Benford's law.

Takeaway

Don’t make up data or statistics. (Obviously, but also because it’s pretty easy to catch you).

More importantly, know that you can look at data and figure out whether to look more closely at it or not.

Hypothesis Testing

We are researching jelly beans and whether they cause acne in teenagers.

The average teen has amount of acne with mean μ and variance σ^2

H_0 (null hypothesis): Jelly beans have no effect on acne (i.e., mean acne is μ for someone who eats jelly beans)

H_A (alternative hypothesis): Jelly beans increase acne (i.e., mean acne is $> \mu$ for someone who eats jelly beans)

Hypothesis Testing

Choose a significance level, say 0.05

Observe 100 jelly bean eating teenagers, measure acne. Sample mean observed: $\bar{x}$

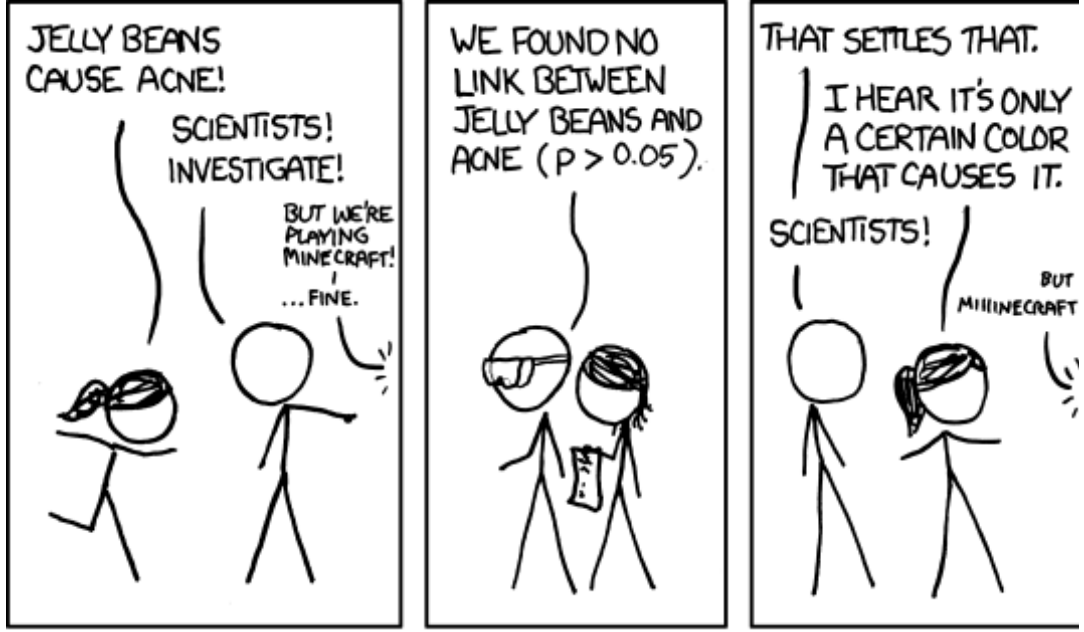
If null hypothesis is true, what is the probability of observing the amount of acne we saw?: $\mathbb{P}(\bar{X} \geq \bar{x}) = \mathbf{0.0162}$

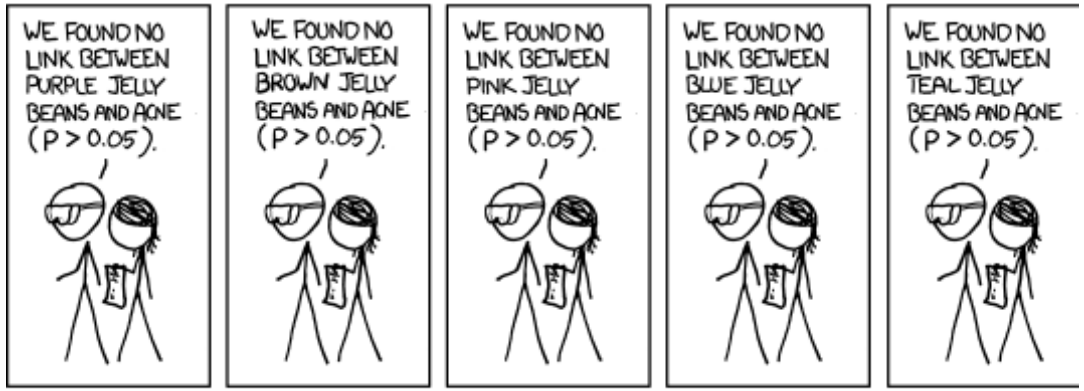
Hypothesis Testing

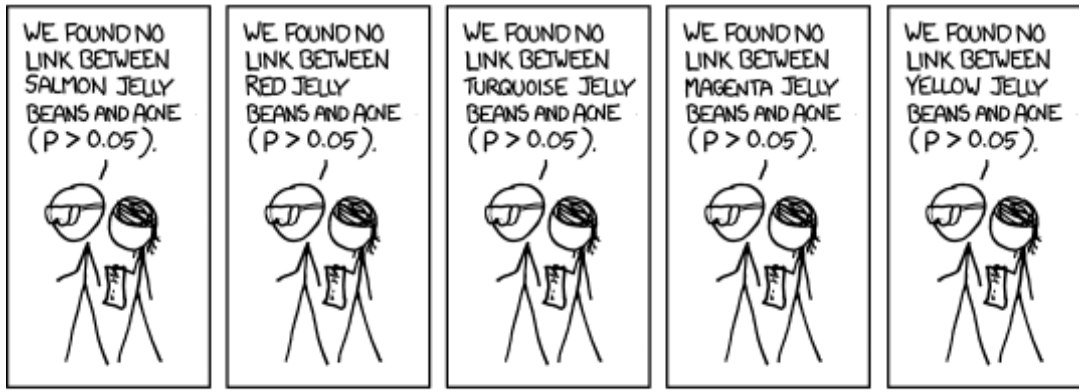
If null hypothesis is true, what is the probability of observing the amount of acne we saw?: $\mathbb{P}(\bar{X} \geq \bar{x}) = \mathbf{0.0162}$

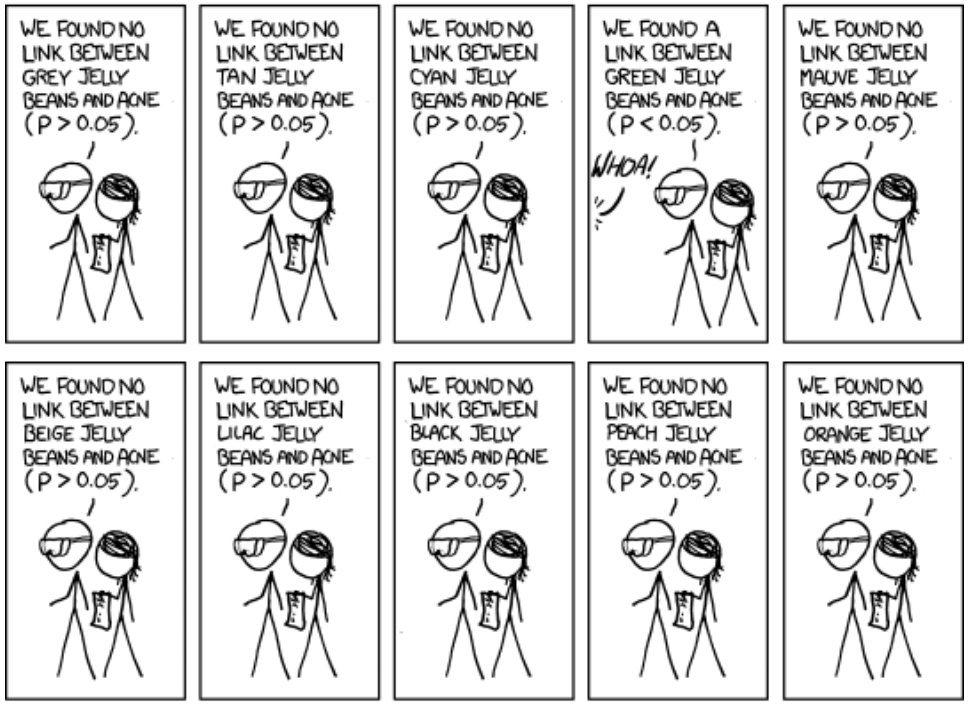
If $p < \mathbf{0.05}$, reject H_0 at the 0.05 significance level (i.e. strong statistical evidence jelly beans cause an increase in acne).

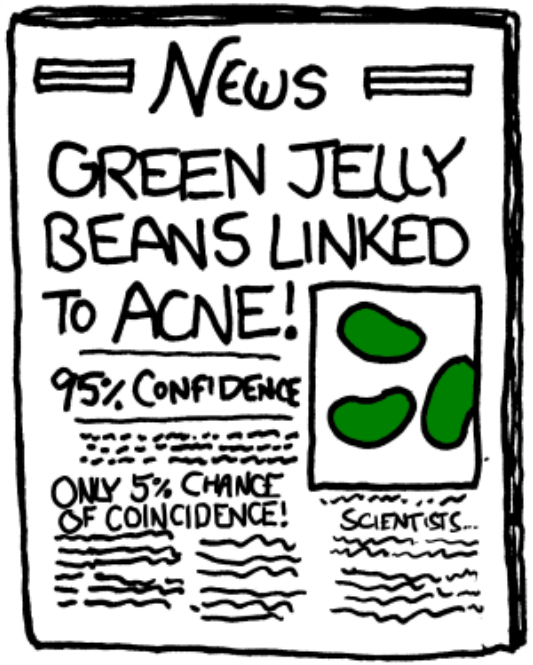
If $p > \mathbf{0.05}$, accept H_0.











https://xkcd.com/882/

p-Hacking

P-value < 0.05: Only a 5% chance of seeing this much acne if jelly beans don’t cause acne.

p-Hacking

P-value < 0.05: Only a 5% chance of seeing this much acne if jelly beans don't cause acne.

What happens if I repeat the experiment 20 times?

$$\mathbb{P}(pval < 0.05 \; at \; least \; 1x) = 1 - (pval > 0.05)^{20} = 1 - 0.95^{20} \approx 0.64$$

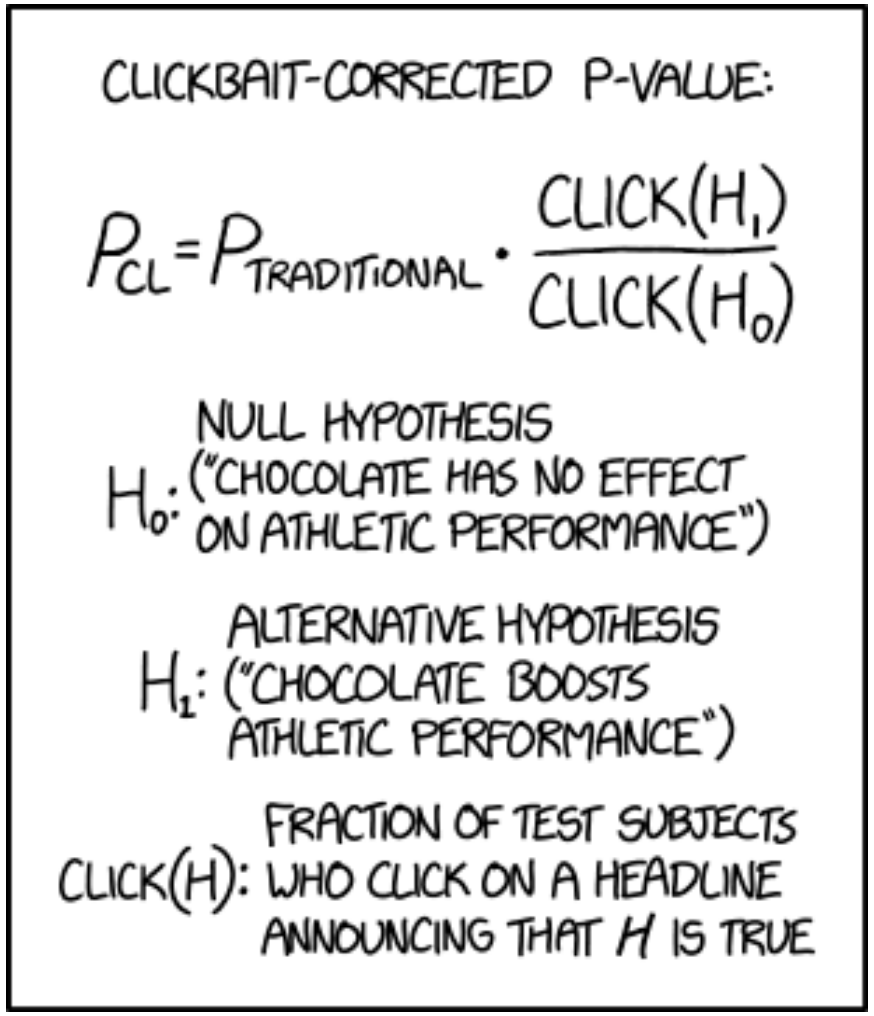
A truly significant result!

p-Hacking

Performing a hypothesis test multiple times in order to get a statistically significant result…

And not reporting the 19 insignificant tests!

Takeaways



Don’t p-hack.

Know that p-hacking is very prevalent in industry.

Leads to better headlines.

Publication bias.

Science is human-motivated, sometimes. Maybe most of the time.

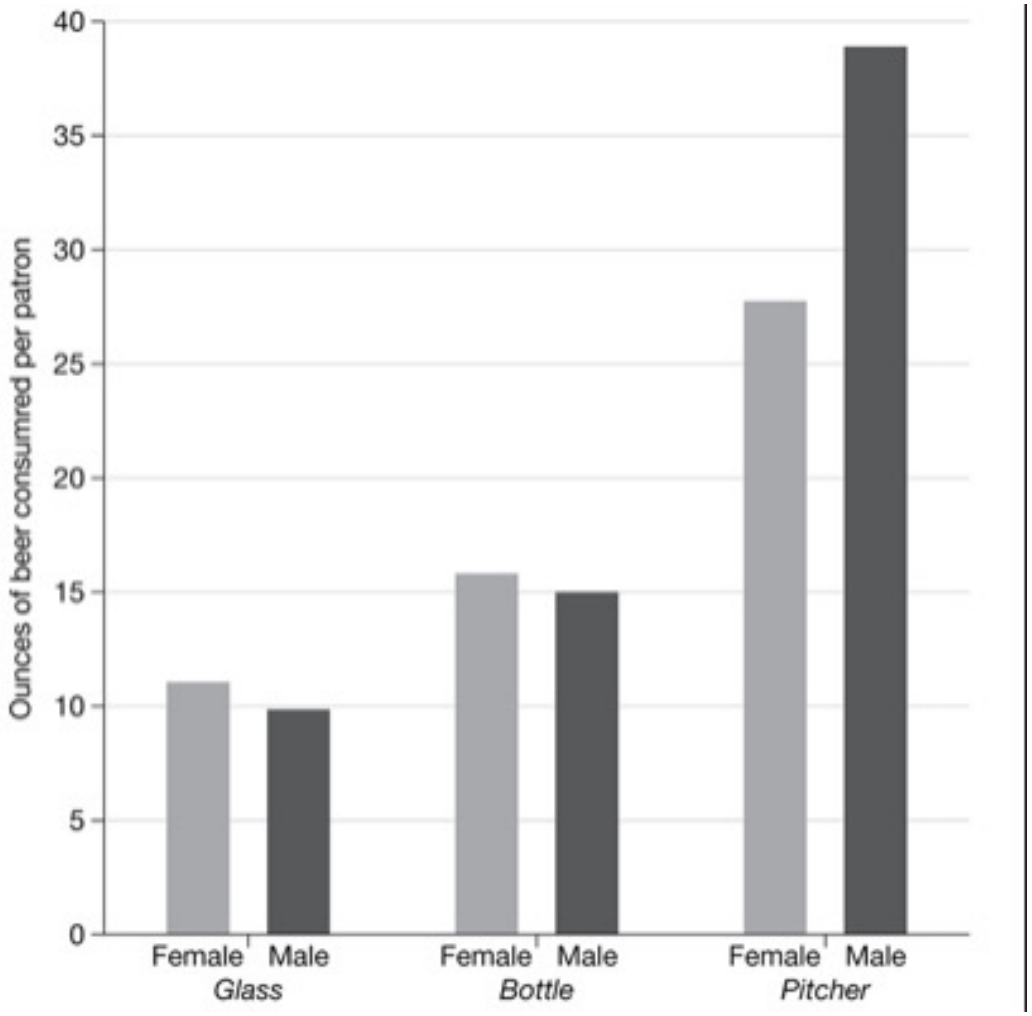
Correlation

"Naturalistic Observations of Beer Drinking among College Students," by Geller et. al

"People drink more when beer is consumed in pitchers" =>

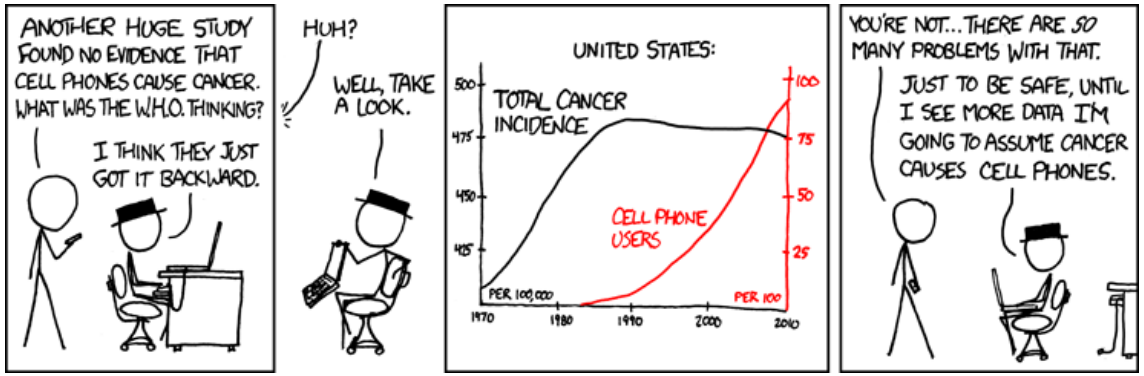
"People drink more because beer is consumed in pitchers." =>

"We should ban pitchers so that students will drink less."



from *Calling Bullshit*, Chapter 4

Correlation



Correlation doesn't sell newspapers.

People **want** to read prescriptive claims.

Graphs often subtly imply causality.

https://xkcd.com/925/

Post hoc fallacy

Causation flows forwards in time. If A happens before B, B cannot cause A.

post hoc, ergo propter hoc (fallacy) – “after this, therefore because of this”

Types of Causation

1. Probabilistic cause
2. Sufficient Cause
3. Necessary Cause

"Time for a quick reality check. Despite the hysteria from the political class and the media, smoking doesn't kill. In fact, 2 out of every three smokers does not die from a smoking related illness and 9 out of ten smokers do not contract lung cancer." – Mike Pence

Omitted variable bias

Consider these headlines (made-up):

“Playing Golf Risk Factor for Cardiac Disease”

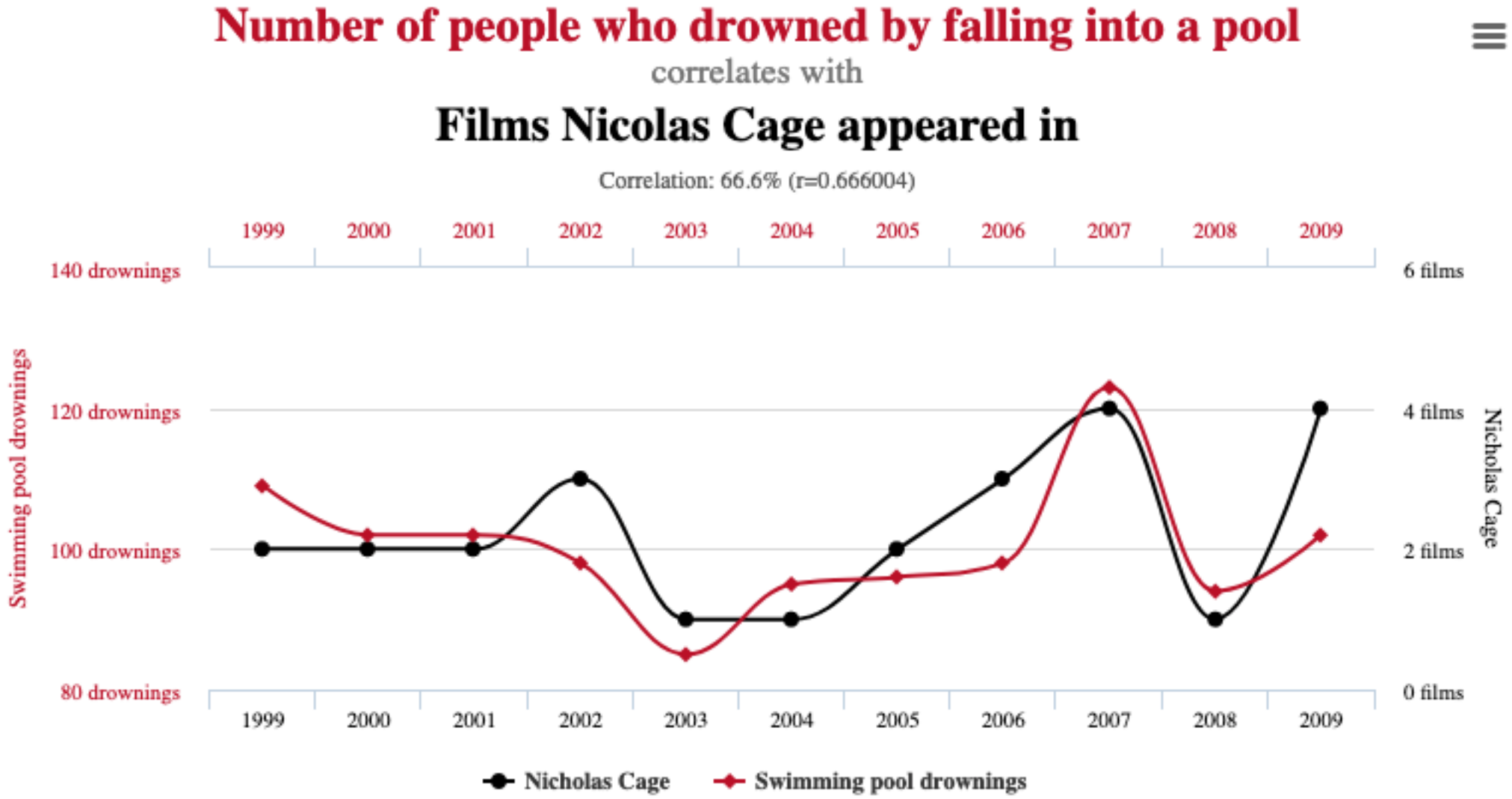
“More Doctors Thought to Contribute To High Levels of Disease in Place”

How do we actually show causation?

When possible, manipulative experiments – try to isolate the hypothesized cause and keep everything else constant.

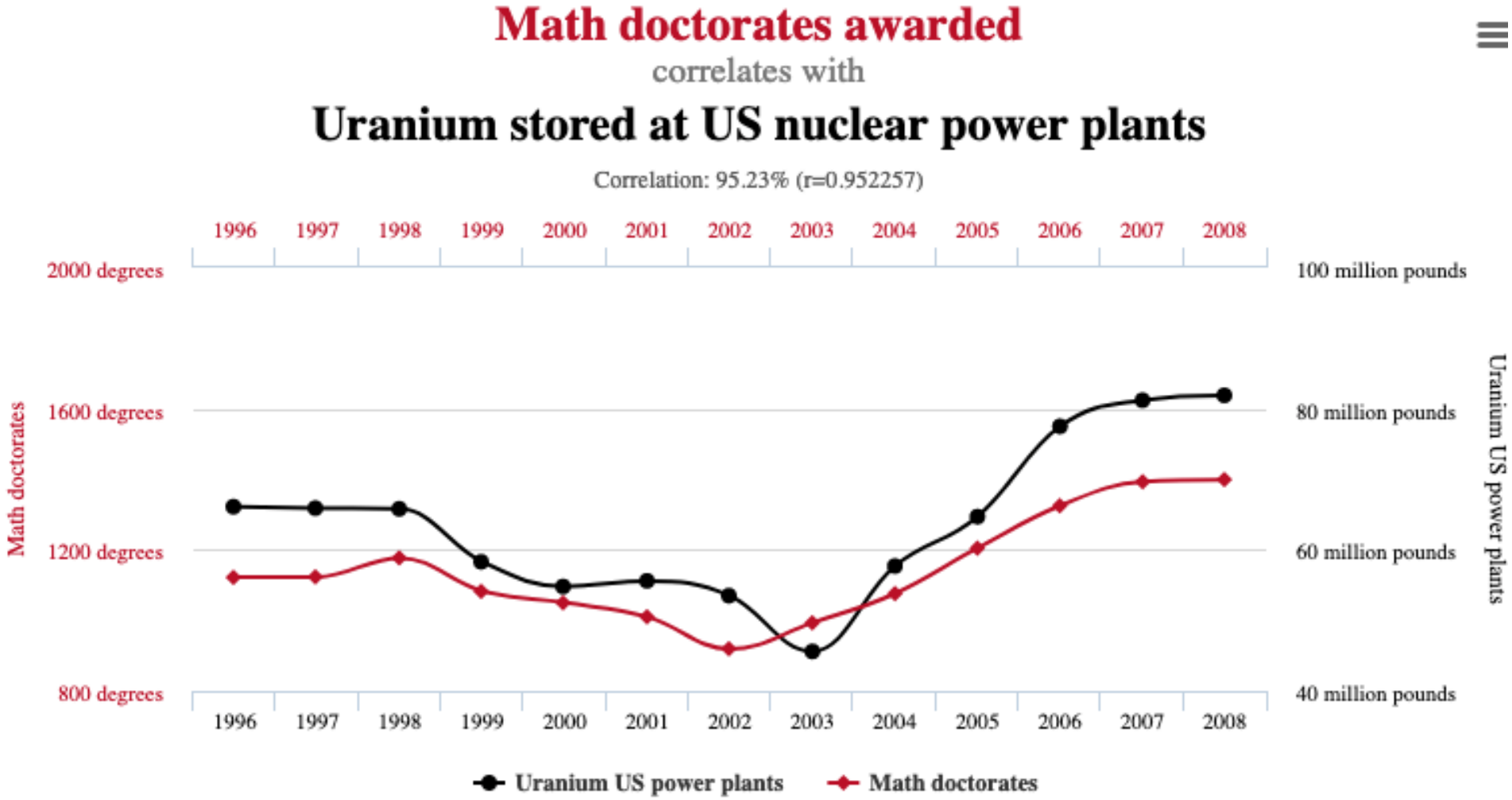
Not always practical.

Spurious Correlations

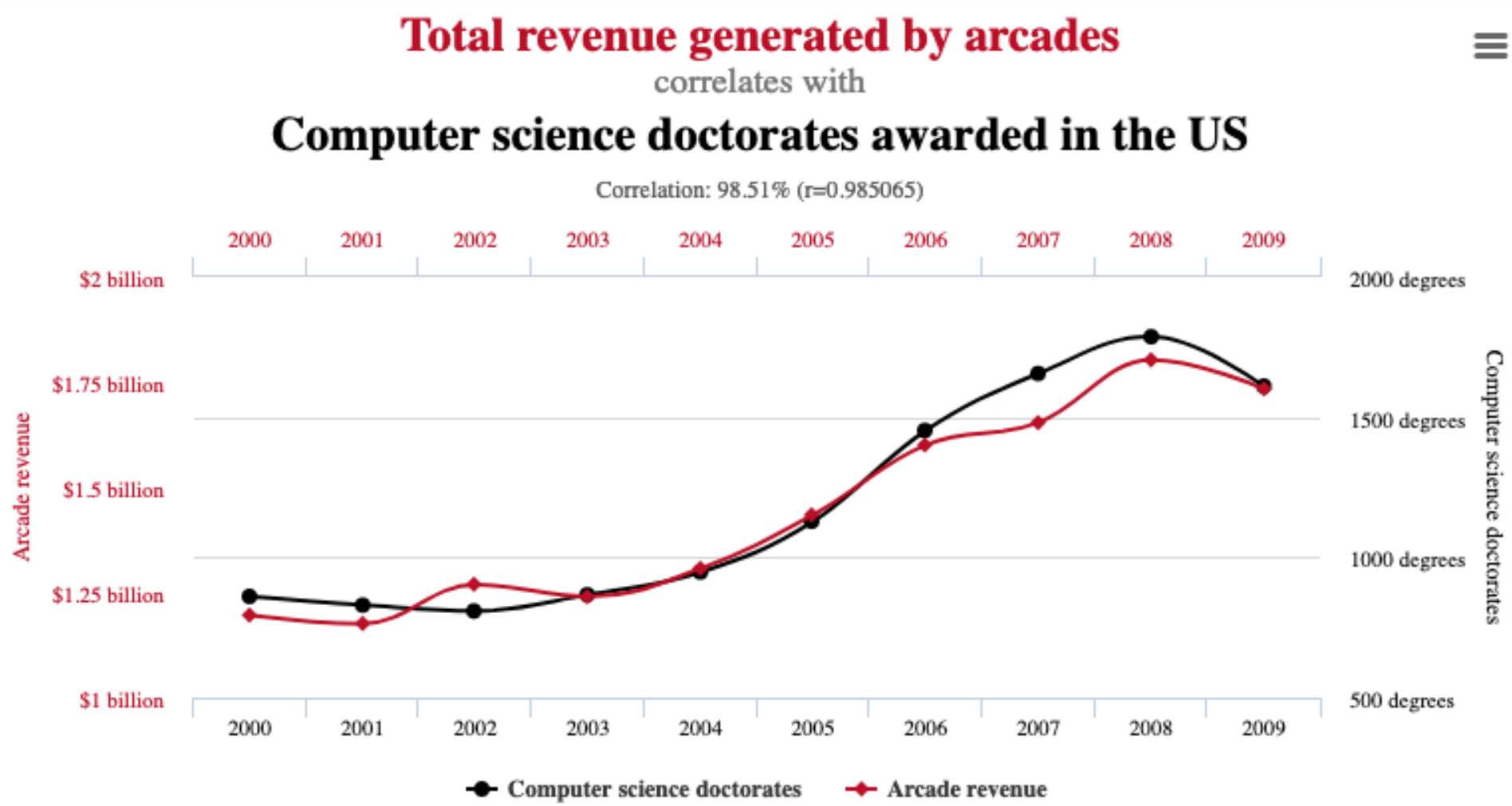


https://www.tylervigen.com/spurious-correlations

Spurious Correlations



Spurious Correlations



Spurious Correlations

What are the chances? Very small. It must mean something, right?

Spurious Correlations

What are the chances? Very small. It must mean something, right?

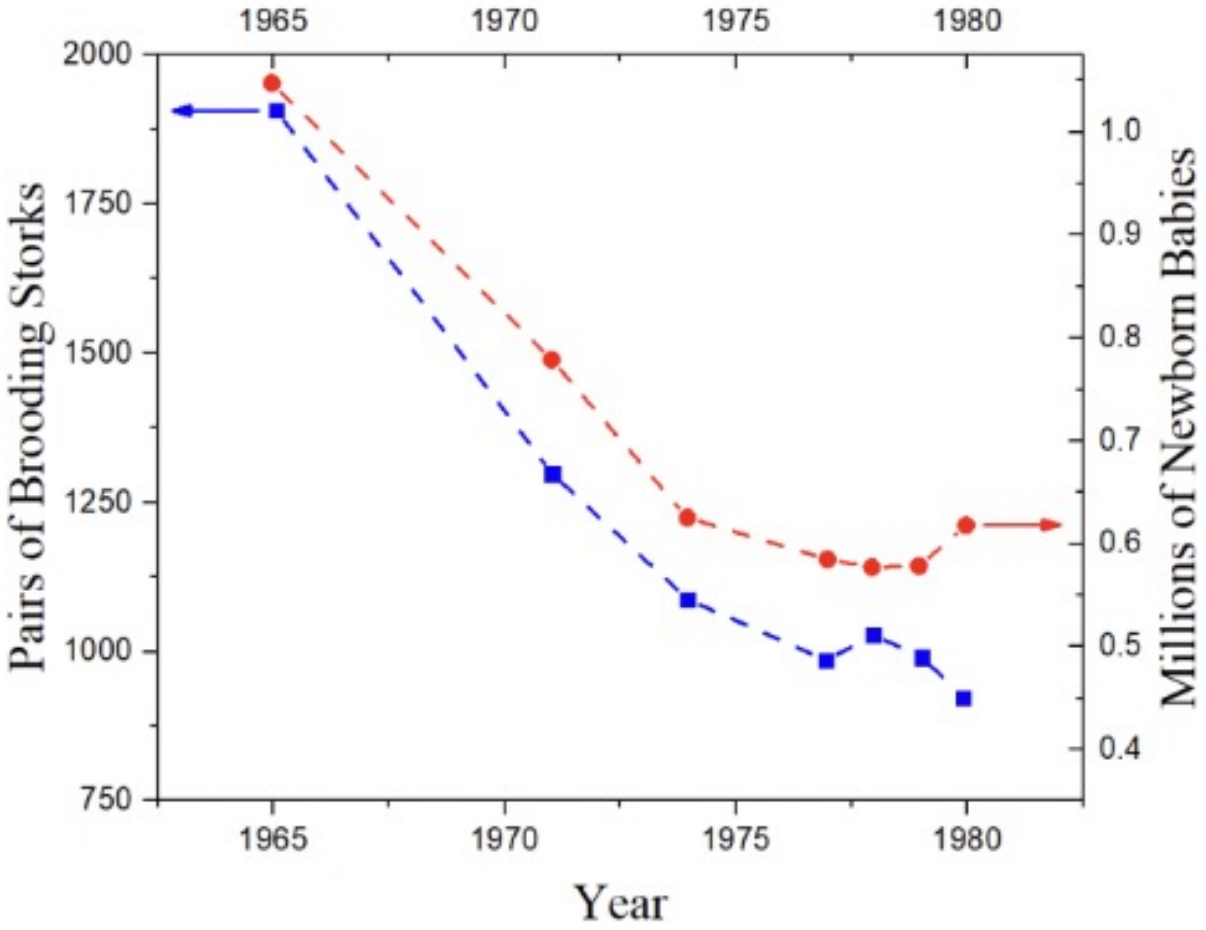
Well, sure. That a lot of things were compared before a correlation was found.

It's less fun in science – when data has a lot of variables, we basically have this exact scenario where just by chance, we can find a correlation.

In fact, super easy to find correlations over time!

Spurious Correlations

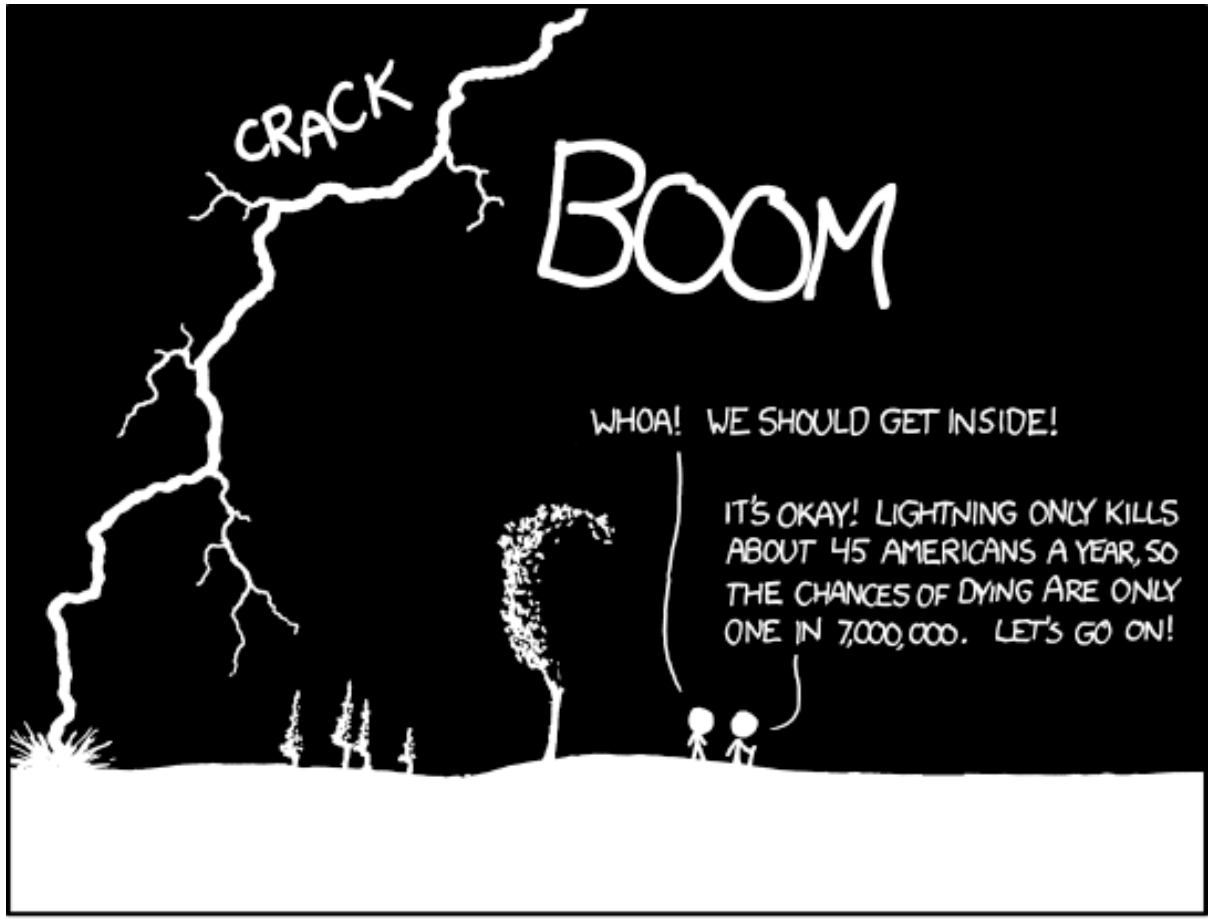
Where Do Babies Come From? By Marco Luis Ferreira Nascimento



Takeaways

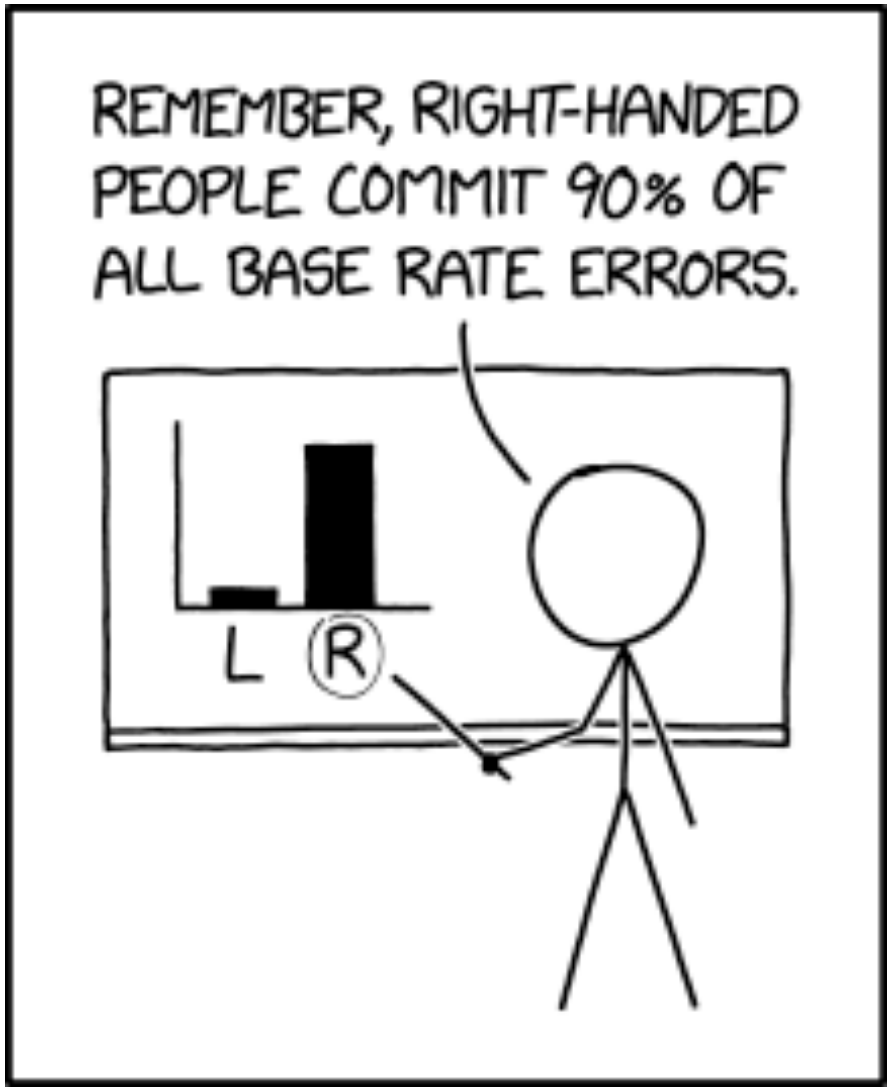
Be wary of people jumping from correlation to causation, and even worse, to prescriptive claims.

Base vs. Conditional Probabilities



https://xkcd.com/795/

Base Rate Fallacy



https://xkcd.com/2476/

Average Class Size

UW says the average class size is 28.

Does that feel true to you?

Is it true?

Probably true!

Let's take a university with 100 students.

5 classes, of sizes: 96, 1, 1, 1, 1.

The average class size is $96 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} = 20$. Sounds great!

Well…

Smaller classes have fewer seats, so on average, we find ourselves in larger classes (a greater proportion of total seats).

What about the average class size *experienced* by students?

$$96 \cdot \frac{96}{100} + 1 \cdot \frac{1}{100} + 1 \cdot \frac{1}{100} + 1 \cdot \frac{1}{100} + 1 \cdot \frac{1}{100} = 92.2$$

Mean vs. Median

George W. Bush administration: 92 million Americans would receive an average tax reduction of $1,083.

Would 92 million Americans be getting a tax cut? Yes.

Would most of those people be getting a tax cut of around $1,000? No. The median tax cut was less than $100.

The median is not sensitive to outliers ☺

Mean vs. Median

Suppose someone has a particular fatal illness. The doctor says that the median improvement of life expectancy on taking a particular medication is 2-3 weeks. The insurance company too, refutes paying for medication based on this statistic.

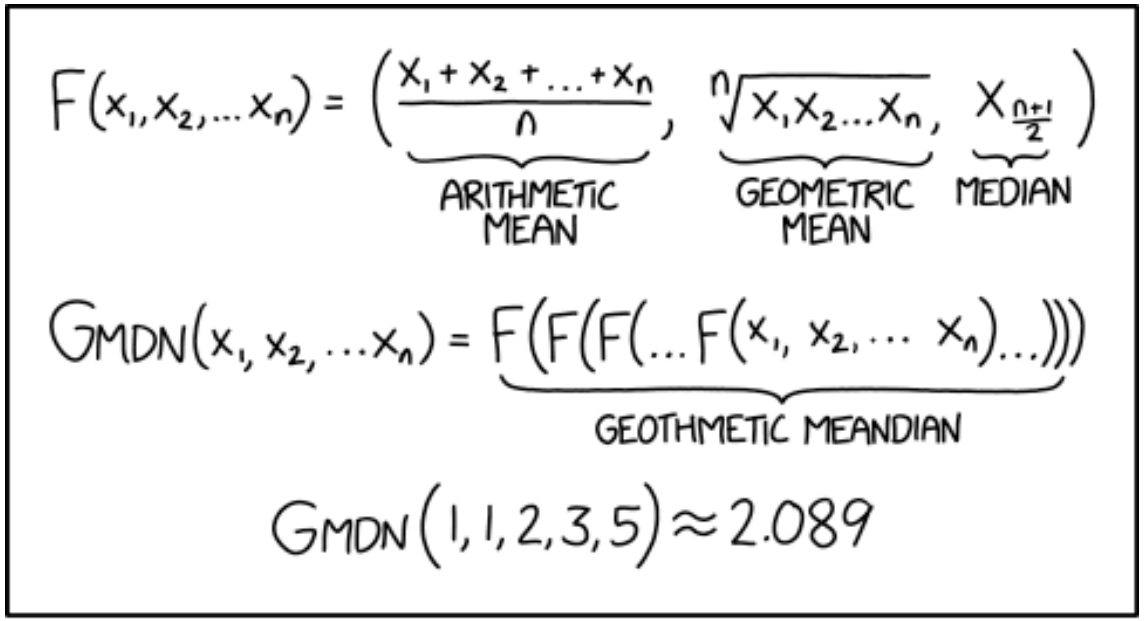
But 30-40% of people fully recover after taking the medication!

The median is not sensitive to outliers ☹

Takeaways

Everyone loves single-shot numbers. Especially averages.

To make yourself look better, use one whose context is not best for the situation. Don't specify how an average was calculated. Nor which was used (mean, median, mode).



STATS TIP: IF YOU AREN'T SURE WHETHER TO USE THE MEAN, MEDIAN, OR GEOMETRIC MEAN, JUST CALCULATE ALL THREE, THEN REPEAT UNTIL IT CONVERGES

https://xkcd.com/2435/

What do you imagine a sequence of 5 flips of a fair coin would look like?

What do you imagine a sequence of 5 flips of a fair coin would look like?

Chances are, what you picture looks more like HTHHT than HHHHH!

Back in 1913…

In the Monte Carlo Casino…

The roulette ball landed on a black number 26 times in a row.

Would it land on black or red on the next spin?

Back in 1913…

In the Monte Carlo Casino…

The roulette ball landed on a black number 26 times in a row.

Would it land on black or red on the next spin?

Gamblers lost millions betting on red!

Gambler's Fallacy

Humans tend to look for balance.

A team has had a losing streak, so they have to have a win soon.

In fact…

> Baseball umpires 5% less likely to call a strike if the previous 2 pitches were strikes.

> US judges 3% more likely to reject asylum if they accepted the previous case

> Loan officers in India up to 23% less likely to approve a loan application if they approved the prior one.

"Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," by Chen, Moscowitz and Shue

Takeaway

A coin has no memory.

Hot hand version: a winning streak.

Sometimes, we fall victim to others though.

College Admissions

In 1973, UC Berkeley was sued for discrimination.

Of all the female students who applied, 35% were admitted.

Of all the male students who applied, 44% were admitted.

What caused this discrimination?

They took a closer look at individual departments… here are the biggest 6 of 85.

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	**4526**	**39%**	**2691**	**45%**	**1835**	**30%**

Legend:

greater percentage of successful applicants than the other gender

greater number of applicants than the other gender

bold - the two 'most applied for' departments for each gender

https://en.wikipedia.org/wiki/Simpson%27s_paradox

What caused this discrimination?

Why? Looking at the rejection rates, women tended to apply to more competitive departments with lower rates of admission even among qualified applicants.

Simpson’s Paradox

A trend appears in several groups of data but reverses/disappears when the groups are combined.

Takeaways

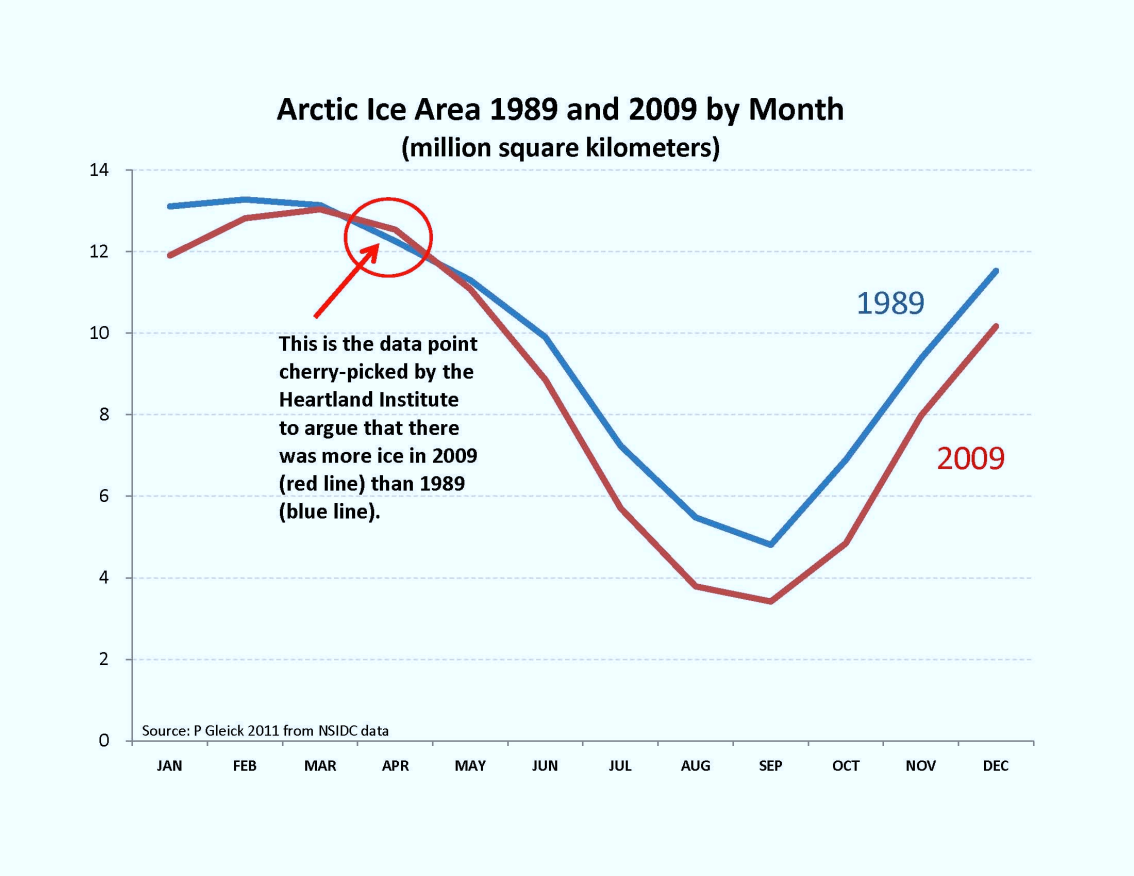
It’s important to think about data as the whole, and as the parts that make it up.

That means different dimensions, too!

Cherry-Picking

"In fact, National Snow and Ice Data Center records show conclusively that in April 2009, Arctic sea ice extent had indeed returned to and surpassed 1989 levels." – Joseph Bast (Heartland Institute, 2011)

Time frame matters!

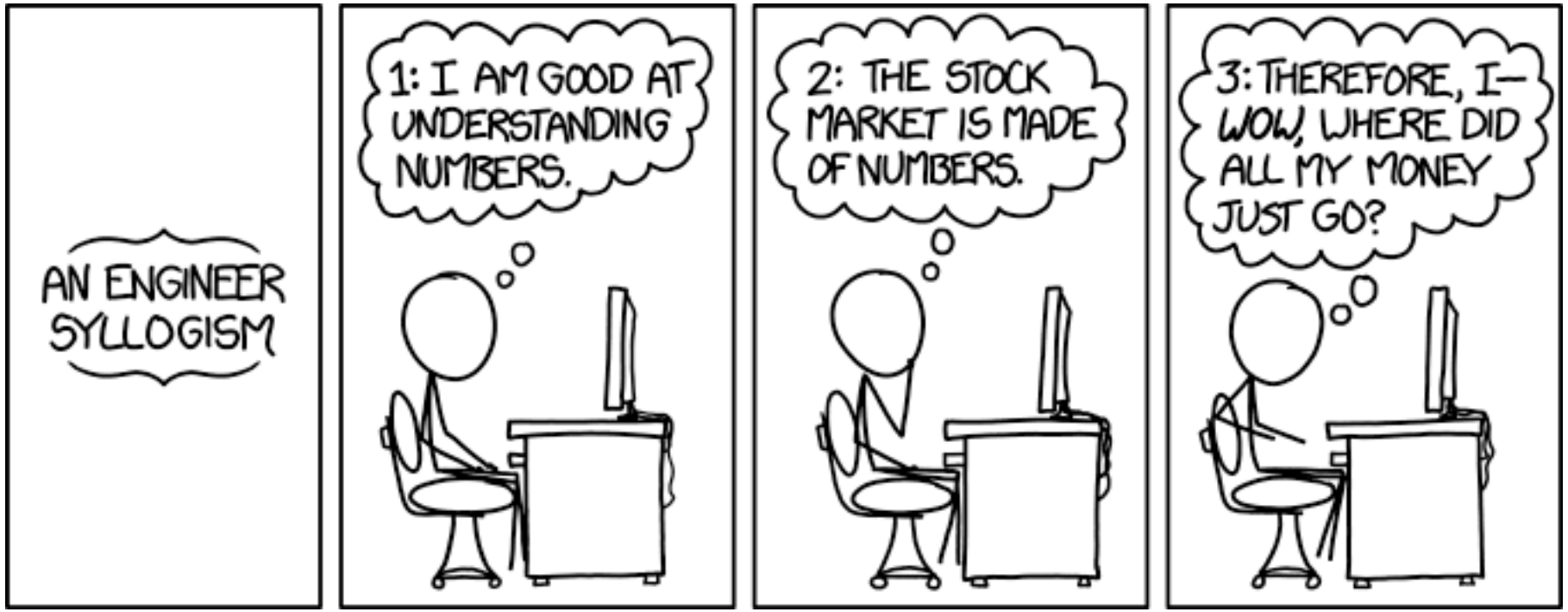


Black Swan Events

2008 Financial Crisis

Risk barometer: Value at Risk model (VaR)

An indicator value that quantified overall risk.



https://xkcd.com/1570/

Black Swan Events

2008 Financial Crisis

Risk barometer: Value at Risk model (VaR)

An indicator value that quantified overall risk.

Underlying risks are not predictable.

False sense of security.

The future doesn’t look like the past.

99% assurance.

Tail Risks

“The greatest risks are never the ones you can see and measure, but the ones you can’t see and therefore can never measure. The ones that seem so far outside the boundary of normal probability that you can’t imagine they could happen in your lifetime—even though, of course, they do happen, more often than you care to realize.” – Joe Nocera

Why so many statistics?

A world of big data.

Descriptive/summary statistics.

- Batting averages.
- Human development index.
- Median/mean income.

Statistics

Statistics seem rooted in math – and math is exact.

Trying to describe complex phenomena with a statistic - not so much.

Understanding Statistics

Often, the problem is not that a statistic is incorrect.

> Context matters.

> Units matter.

> Relative numbers matter.

> Research: choose what to study and how to study it

Truths can compete: there may be multiple interpretations of facts.