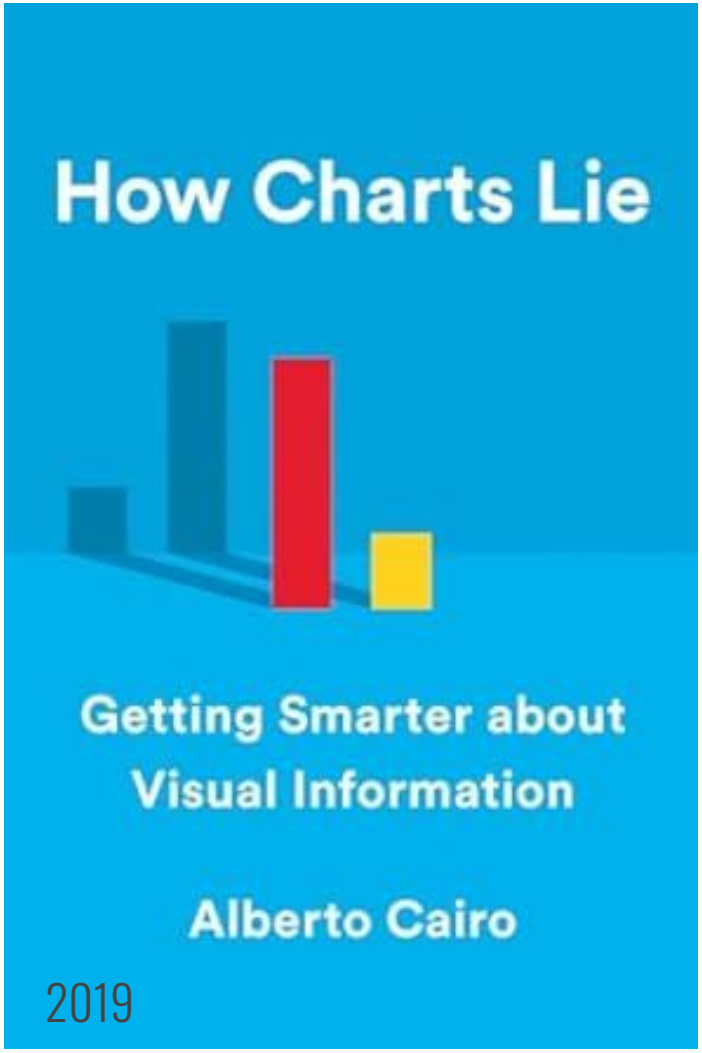
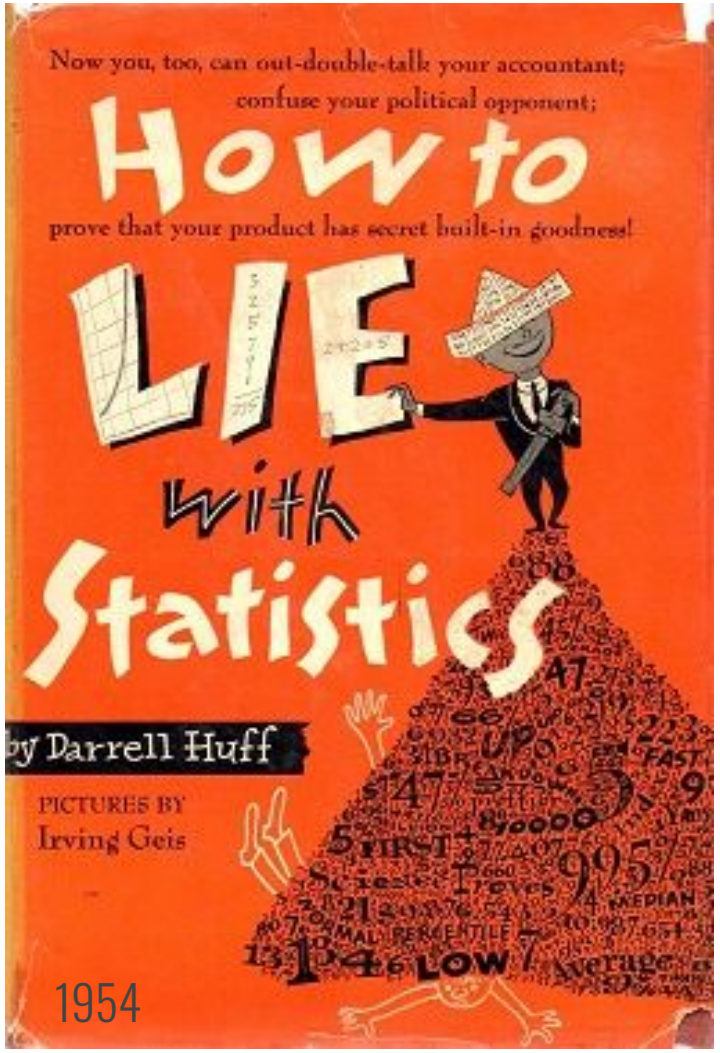
How to Be Wrong

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

Lecture concept and slides built off of Jason Fleischer's lecture of the same name


How Charts Lie
Getting Smarter about Visual Information
Alberto Cairo
2019


Now you, too, can out-double-talk your accountant; confuse your political opponent;
How to
prove that your product has secret built-in goodness!
LIE with Statistics
by Darrell Huff
PICTURES BY Irving Geis
1954

- **A** Thought about it and have a thought
- **B** Thought about it and have no thought
- **C** I'm confused

Top Cities For BBQ in the U.S.

An analysis of TripAdvisor restaurant reviews by chefspencil.com

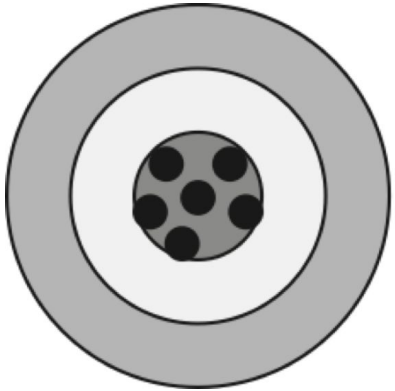
Worst Cities For BBQ in the U.S.

An analysis of TripAdvisor restaurant reviews by chefspencil.com

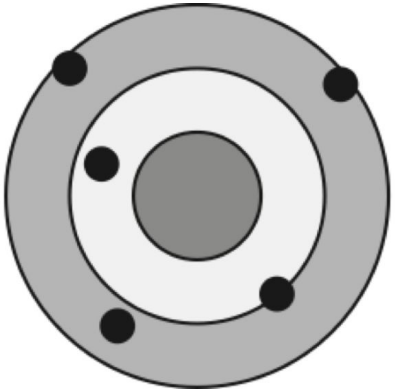
https://www.wspa.com/news/tripadvisor-sets-the-record-straight-after-bbq-ranking-left-kansas-city-out-of-top-30/

Errors of measurement

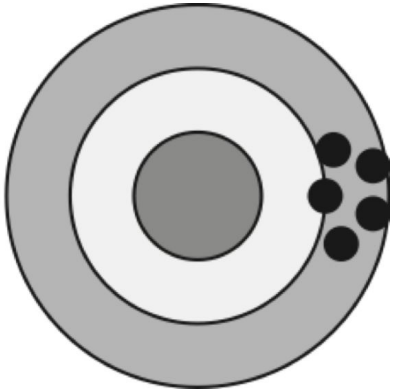
Accurate and precise



Accurate but not precise



Precise but not accurate



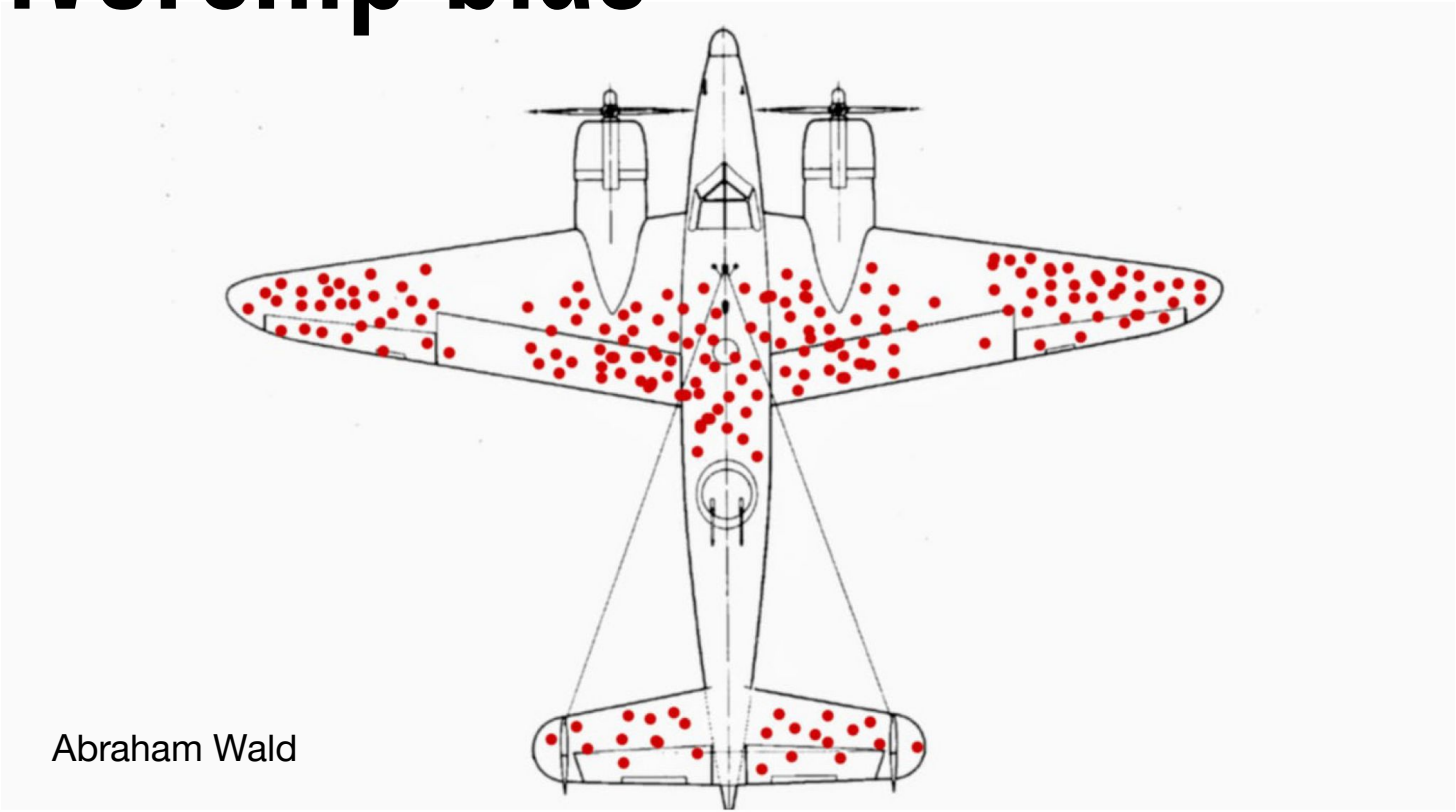
Proxy Measurements

For when we can’t measure directly what we care about, but we can measure something related

- Unemployment rate → our general economy
- Gross domestic product → standard of living
- BMI → health
- Your suggestions?
- Grades → learning

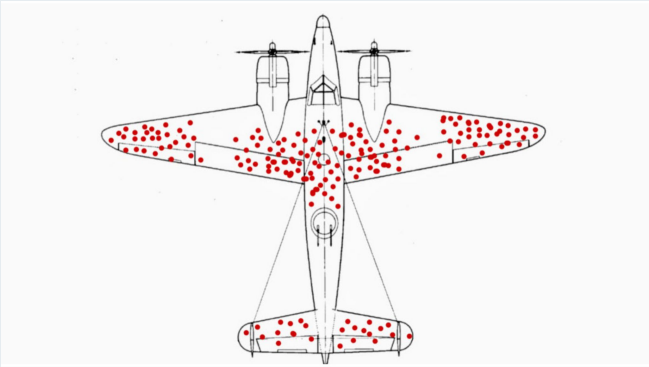
Do we understand what we are measuring?

Survivorship bias

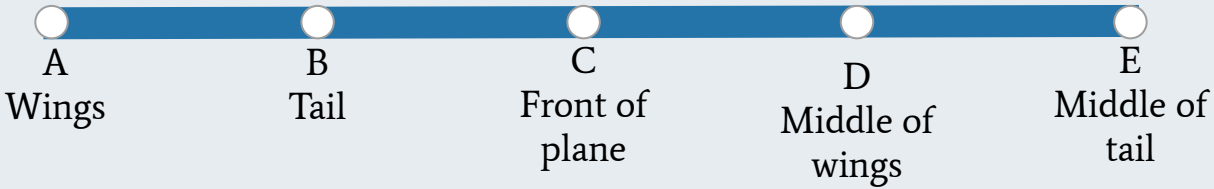


Abraham Wald

Survivorship Bias



Where would you put the armor?



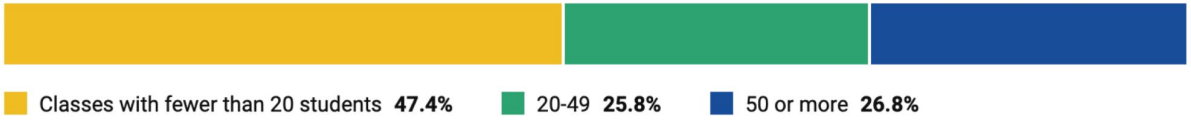
Are we measuring what's relevant?

- **A** This reflects my experience
- **B** This kind of reflects my experience
- **C** This does not reflect my experience

Academic Life at University of California--San Diego

The student-faculty ratio at University of California--San Diego is 19:1, and the school has 47.4% of its classes with fewer than 20 students. The most popular majors at University of California--San Diego include: Biology, General; Mathematics; Economics; International/Global Studies; and Computer Science. The average freshman retention rate, an indicator of student satisfaction, is 94%.

Class Sizes



Classes with fewer than 20 students **47.4%** 20-49 **25.8%** 50 or more **26.8%**

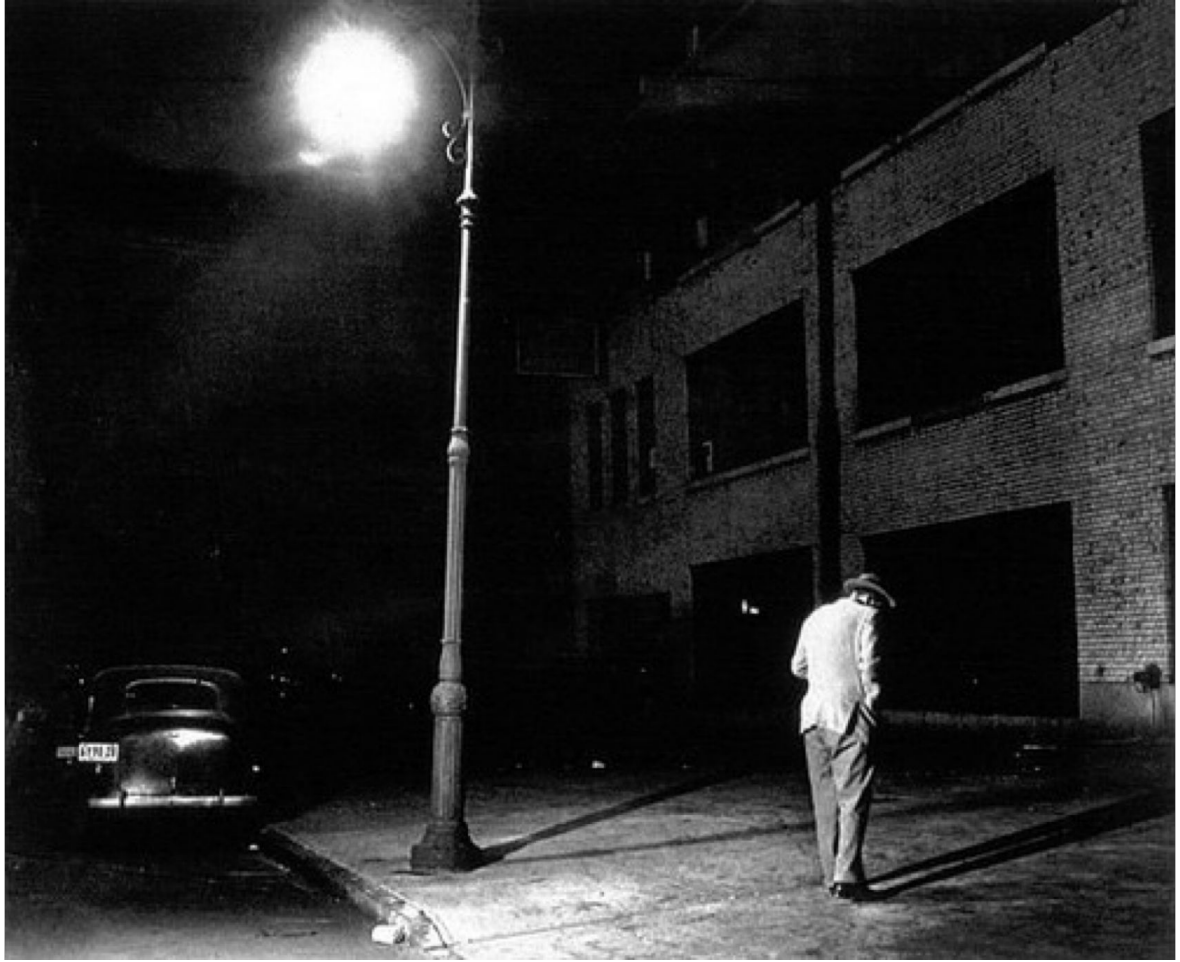
Student-faculty ratio	**19:1**
4-year graduation rate	**65%**

US News & World Report rankings

		% of classes with this many students	Cumulative %	Fraction of classes with this many students * min number of students in that class type	% of students in these classes (normalized version of column to the left)	
	2-9 students:	12%	12%	0.24	0.67%	
	10-19 students	32%	44%	3.2	8.95%	
Median class size as experienced by faculty	20-29 students:	14%	58%	2.8	7.83%	
	30-39 students:	8%	66%	2.4	6.72%	
	40-49 students:	4%	70%	1.6	4.48%	
	50-99 students:	11%	81%	5.5	15.39%	
Median class size as experienced by students	Over 100 students:	20%	101%	20	55.96%	
			Sum:	35.74		
	Data from https://www.collegedata.com/college/University-of-California-San-Diego/					

The Lamppost Problem

aka. Streetlight Effect
aka. The Drunkard's Search
aka. Picking a crap proxy



Goodhart's law

Once a metric becomes a goal, you're hosed

Volkswagen emissions scandal

26 languages

Article Talk | Read Edit View history Tools

From Wikipedia, the free encyclopedia

"Dieselgate" and "Emissionsgate" redirect here. For other diesel emissions scandals, see Diesel emissions scandal.

The **Volkswagen emissions scandal**, sometimes known as **Dieselgate**[23][24] or **Emissionsgate**,[25][24] began in September 2015, when the United States Environmental Protection Agency (EPA) issued a notice of violation of the Clean Air Act to German automaker Volkswagen Group.[26] The agency had found that Volkswagen had intentionally programmed turbocharged direct injection (TDI) diesel engines to activate their emissions controls only during laboratory emissions testing, which caused the vehicles' NO_x output to meet US standards during regulatory testing. However, the vehicles emitted up to 40 times more NO_x in real-world driving.[27] Volkswagen deployed this software in about 11 million cars worldwide, including 500,000 in the United States, in model years 2009 through 2015.[28][29][30][31]

Volkswagen emissions scandal

A 2010 Volkswagen Golf TDI displaying "Clean Diesel" at the Detroit Auto Show

Search engine optimization

65 languages

Article Talk | Read View source View history Tools

From Wikipedia, the free encyclopedia

"SEO" redirects here. For other uses, see Seo.

Search engine optimization (**SEO**) is the process of improving the quality and quantity of website traffic to a website or a web page from search engines.[1][2] SEO targets unpaid traffic (known as "natural" or "organic" results) rather than direct traffic or paid traffic. Unpaid traffic may originate from different kinds of searches, including image search, video search, academic search,[3] news search, and industry-specific vertical search engines.

As an Internet marketing strategy, SEO considers how search engines work, the computer-programmed algorithms that dictate search engine behavior, what people search for, the actual search terms or keywords typed into search engines, and which search engines are preferred by their targeted audience. SEO is performed because a website will receive more visitors from a search engine when websites rank higher on the search engine results page (SERP). These visitors can then potentially be converted into customers.[4]

Part of a series on
Internet marketing
Search engine optimization
Local search engine optimisation
Social media marketing
Email marketing
Referral marketing
Content marketing
Native advertising
Search engine marketing
Pay-per-click
Cost per impression
Search analytics
Web analytics
Display advertising
Ad blocking
Contextual advertising

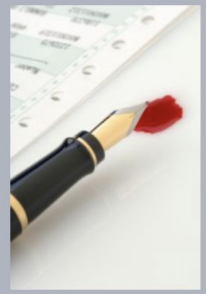
Errors of assumptions around tools

X
Excel

EuSpRIG
European Spreadsheet Risks Interest Group

EuSpRIG Horror Stories

Spreadsheet mistakes - news stories



Public reports of spreadsheet errors have been sought out on behalf of EuSpRIG by Patrick O'Beirne of Systems Modelling for many years. There are very many reports of spreadsheet related errors and they seem to appear in the global media at a fairly consistent rate.

These stories illustrate common problems that occur with the uncontrolled use of spreadsheets. In many cases, we identify the area of risk involved and then say how we think the problem might have been avoided.

Stories are identified by those who kindly collated and sorted them:

POB: Patrick O'Beirne, Eusprig chair

FH: Felienne Hermans (winner of the 2011 David Chadwick student prize and now an assistant professor at Delft University of Technology).

NS: Tie Cheng, a EuSpRIG committee member.

MPC: Mary Pat Campbell, an actuary, trainer, and a member of the EuSpRIG Discussion group.

Identifier:	POB2001
Title:	Data not controlled, 16000 UK Covid-19 test results lost for a week
Source:	https://www.bbc.co.uk/news/technology-54423988
Release Date:	08 October 2020
Risk:	Lives put at risk because the contact-tracing process had been delayed
Discrepancy:	16,000 test cases in a week

Excel: Why using Microsoft's tool caused Covid-19 results to be lost
"The badly thought-out use of Microsoft's Excel software was the reason nearly 16,000 coronavirus cases went unreported in England. [The labs] filed their [result logs] results in the form of text-based lists - known as CSV files - without issue. PHE had set up an automatic process to pull this data together into Excel templates so that it could then be uploaded to a central system. The problem is that [Public Health England] PHE's own developers picked an old file format to do this - known as XLS. As a consequence, each template could handle only about 65,000 rows of data rather than the one million-plus rows that Excel is actually capable of. And since each test result created several rows of data, in practice it meant that each template was limited to about 1,400 cases. When that total was reached, further cases were simply left off. To handle the problem, PHE is now

Ads by Google
Stop seeing this ad Why this ad?

MICROSOFT REPORT SCIENCE

Scientists rename human genes to stop Microsoft Excel from misreading them as dates

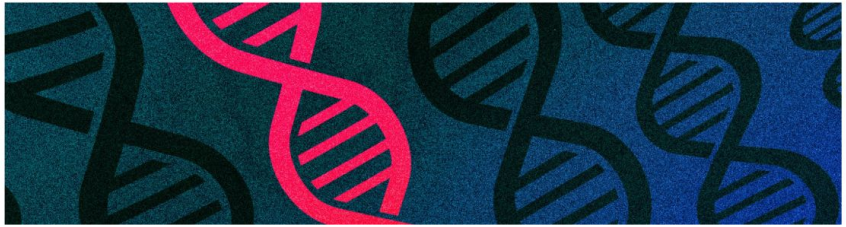
99

Sometimes it's easier to rewrite genetics than update Excel

By James Vincent | Aug 6, 2020, 8:44am EDT

Listen to this article

SHARE



https://www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misreading-dates

What about when Excel is used as intended?

15k spreadsheets
97M cells
20M formulas

Enron's Spreadsheets and Related Emails: A Dataset and Analysis

Felienne Hermans
Delft University of Technology
Mekelweg 4
2628 CD Delft, the Netherlands
f.f.j.hermans@tudelft.nl

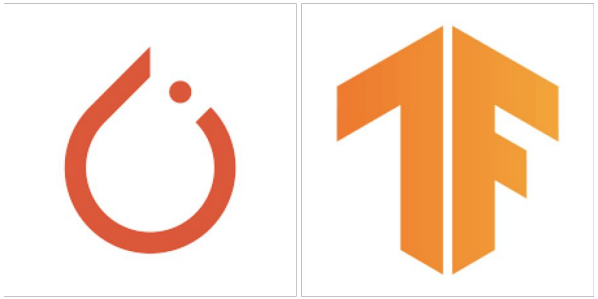
Emerson Murphy-Hill
North Carolina State University
890 Oval Drive
Raleigh, North Carolina, USA
emerson@csc.ncsu.edu

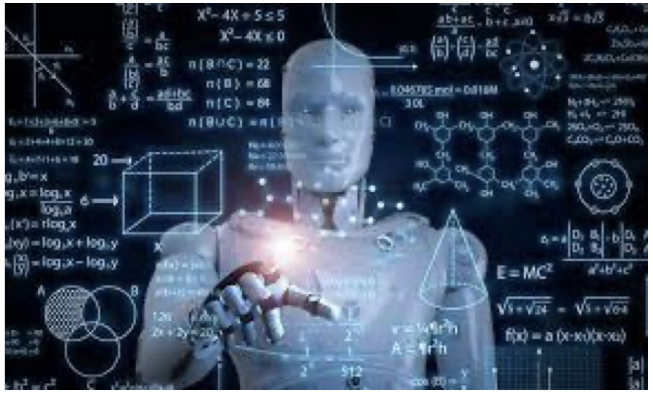
TABLE III
SPREADSHEETS CONTAINING EXCEL ERRORS IN THE ENRON SET

Error type	Spreadsheets	Formulas	Unique Ones
#DIV/0!	580	76,656	4,779
#N/A	635	948,194	6,842
#NAME?	297	33,9365	29,422
#NUM!	52	4,087	178
#REF!	931	18,3014	6824
#VALUE!	423	11,1024	1751
Total	2,205	1,662,340	49,796

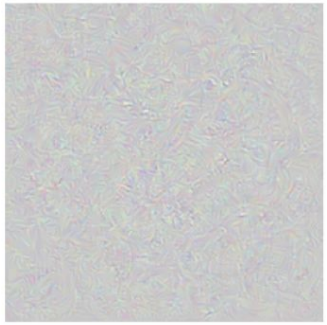
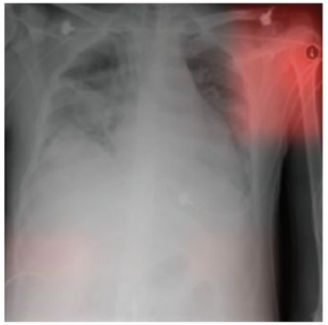
24% of spreadsheets with formulas had errors

Excel





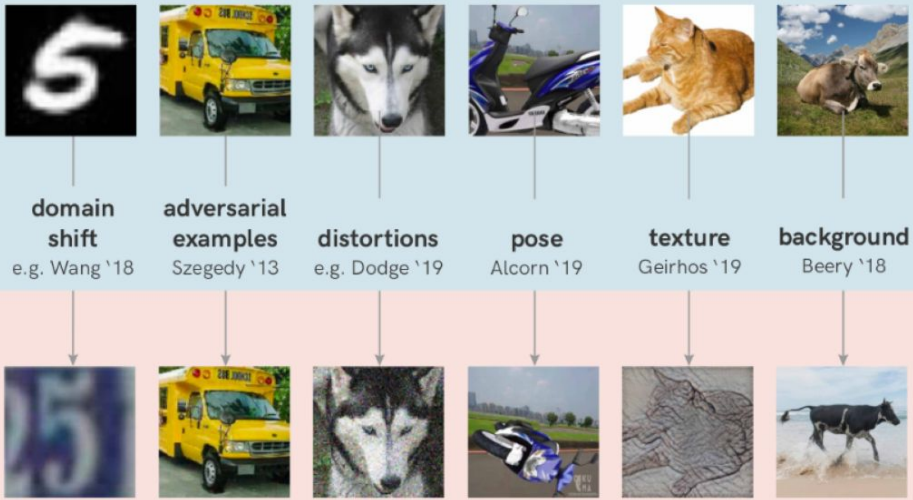
Learning the irrelevant
ML loves the shortcut

				**Article:** Super Bowl 50 **Paragraph:** *"Peython Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had a jersey number 37 in Champ Bowl XXXIV."* **Question:** *"What is the name of the quarterback who was 38 in Super Bowl XXXIII?"* **Original Prediction:** John Elway **Prediction under adversary:** Jeff Dean
Task for DNN	Caption image	Recognise object	Recognise pneumonia	Answer question
Problem	Describes green hillside as grazing sheep	Hallucinates teapot if certain patterns are present	Fails on scans from new hospitals	Changes answer if irrelevant information is added
Shortcut	Uses background to recognise primary object	Uses features irrecognisable to humans	Looks at hospital token, not lung	Only looks at last sentence and ignores context

https://thegradient.pub/shortcuts-neural-networks-love-to-cheat/

same category for humans
but not for DNNs (intended generalisation)

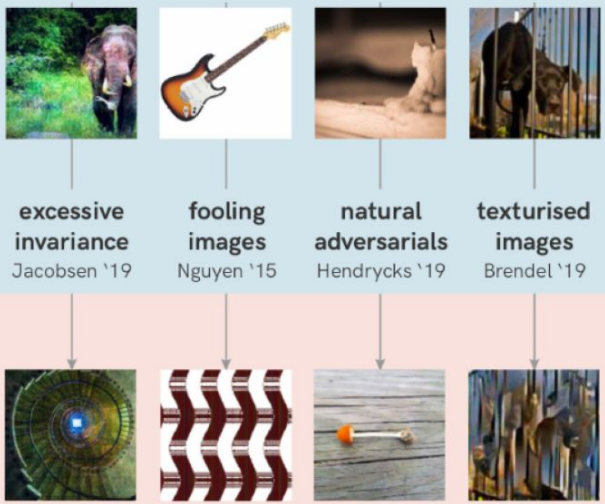
i.i.d.



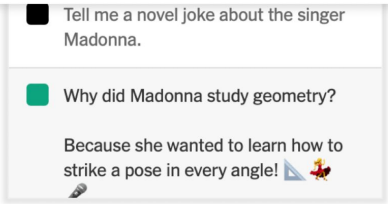
domain shift e.g. Wang '18
adversarial examples Szegedy '13
distortions e.g. Dodge '19
pose Alcorn '19
texture Geirhos '19
background Beery '18

o.o.d.

same category for DNNs
but not for humans (unintended generalisation)



excessive invariance Jacobsen '19
fooling images Nguyen '15
natural adversarials Hendrycks '19
texturised images Brendel '19



10 Ways GPT-4 Is Impressive but Still Flawed

OpenAI has upgraded the technology that powers its online chatbot in notable ways. It's more accurate, but it still makes things up.

By Cade Metz and Keith Collins

Cade Metz asked experts to use GPT-4, and Keith Collins visualized the answers that the artificial intelligence generated.

March 14, 2023

https://www.nytimes.com/2023/03/14/technology/openai-new-gpt4.html

Errors around what we get to know

Preregistration and publishing negative results

"Data available upon reasonable request"

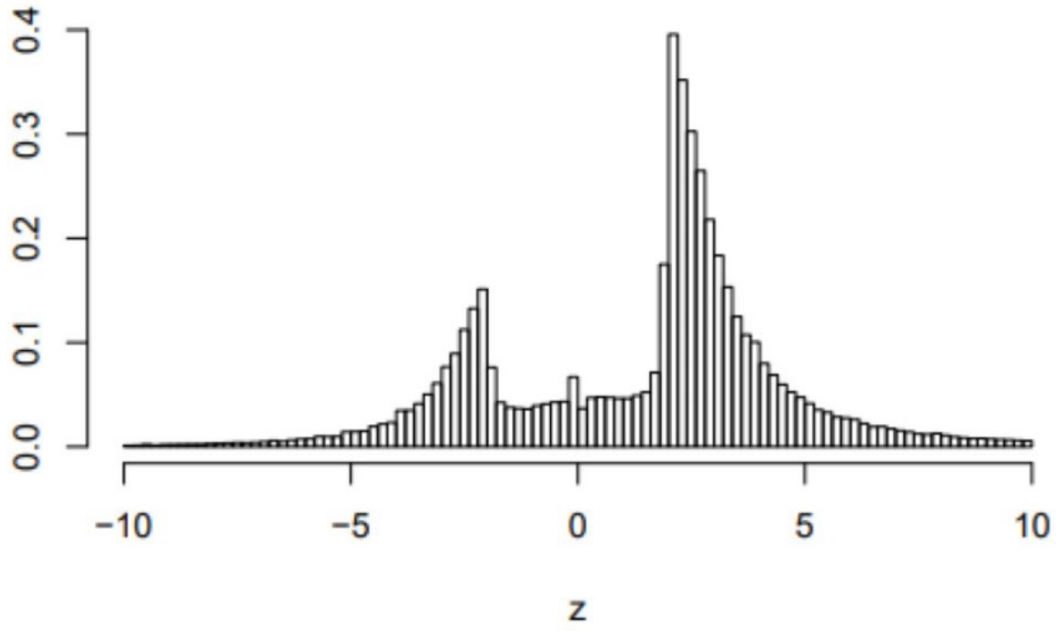
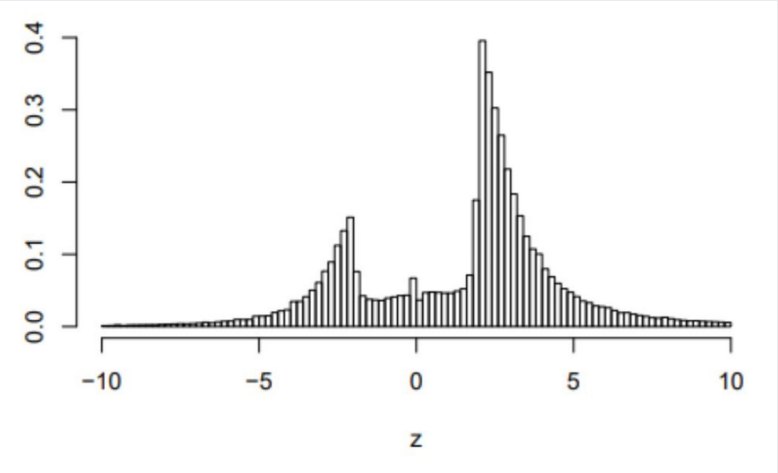


Figure 1: The distribution of more than one million z-values from Medline (1976–2019).

p-values in published research



What explains this distribution?

A
I have an idea

B
I'm not sure

C
I'm confused

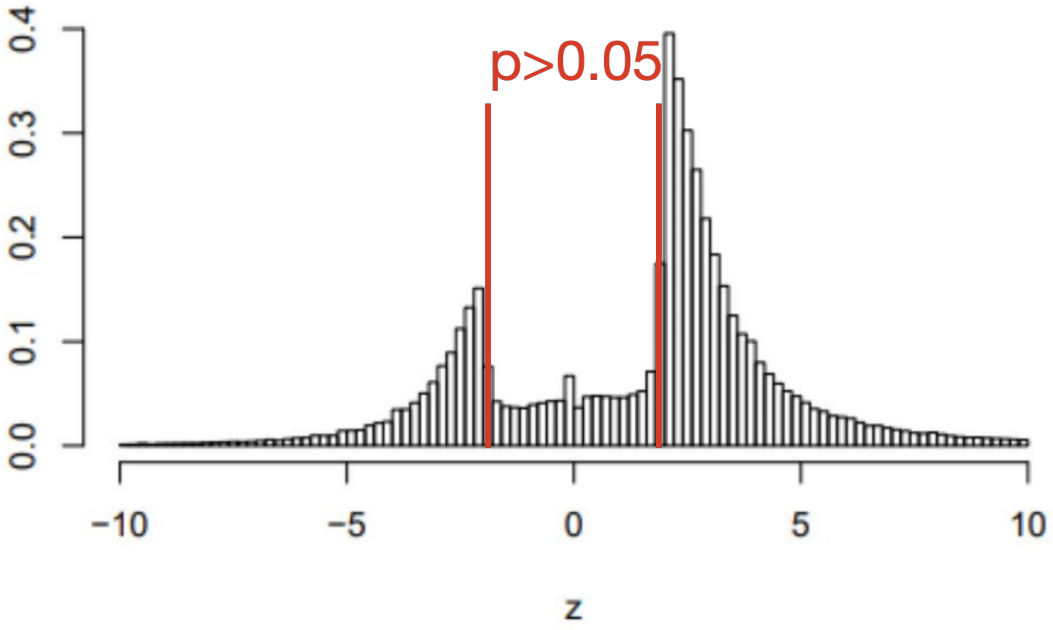


Figure 1: The distribution of more than one million z-values from Medline (1976–2019).

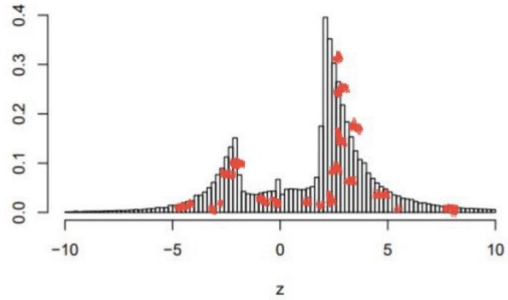
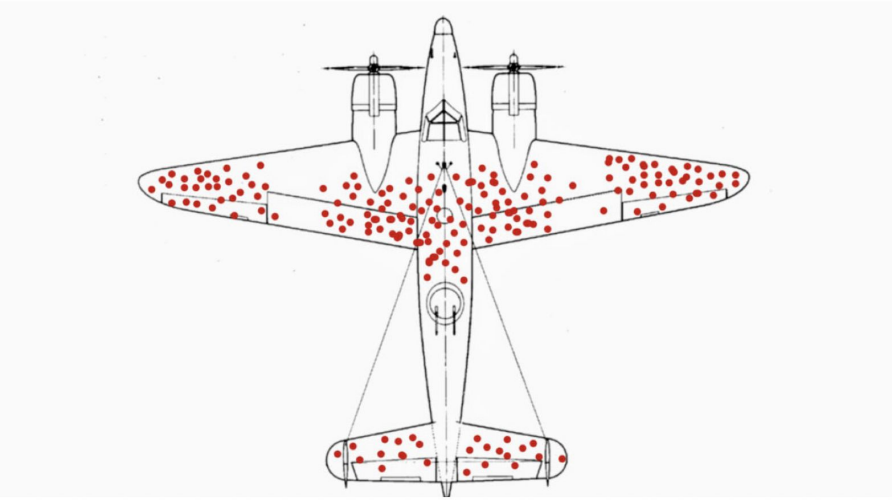



Figure 1: The distribution of more than one million z-values from Medline (1976–2019).