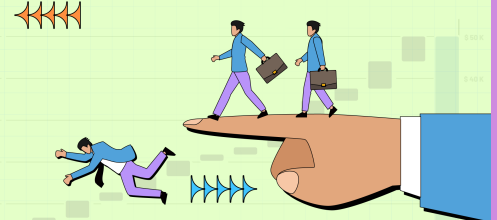


## Misleading Statistics in the Media: Examples and How to Spot Them



### PART 1: How false news can spread

1. Why should we be more careful today about information and data displayed in the news?
2. Oral comprehension.

Watch TED-Ed's video about how false news can spread (00:00-03:02) and answer the following questions.



Read the questions before watching the video so you know what to expect.



Right or wrong? You may have to justify your answer.

1. Decades ago there were fewer newspapers so the information was less trustworthy: **Wrong.** In previous decades most media with global reach consisted of several major newspapers and networks which had the resources to gather information directly.
2. Aggregated sources of news like the Associated Press (AP) are a safe way to avoid the circulation of misinformation: **Wrong.** The speed with which information spreads now has created the ideal conditions for "circular reporting," when publication A publishes misinformation, publication B reprints it, and publication A then cites B as the source for the information.
3. Since information travels fast and everybody can participate, false data are easily corrected: **Wrong.** The pseudo-scientific article linking vaccination and autism has repeatedly been discredited by the scientific community but the belief persists. It is also a form of problematic circular reporting when multiple publications report on the same initial piece of false information which then appears to another author as having been verified by multiple sources. Moreover an unverified fact in a wiki page can make its way into a published article that may later be added as a citation for the very same wiki information.
4. Children that should have been vaccinated were contaminated: **Right.** Deliberately unvaccinated children are now contracting contagious diseases that had been virtually eradicated in the US.
5. The video suggests serious journals should not make jokes: **Wrong.** A joke article in the British Medical Journal has been referenced in serious science publications over 400 times. But the problem is not the joke, it is the fact outlets pick it up not in on the joke.
6. In terms of information, new technologies have had positive effects: **Right.** User-generated content may entail circular reporting, but recent advances in communication technologies have had immeasurable benefits, such as breaking down the barriers between information and people.

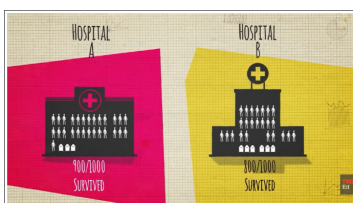
Answer the following question through complete sentences.

7. Why can it be said that Mark Twain's quotation proves the precise point it is making?

The quote usually attributed to Mark Twain says that "A lie can travel halfway around the world while the truth is putting on its shoes." Since there is reason to doubt that Mark Twain ever said this at all, the attribution of the quote is an example of misinformation that "travelled halfway around the world."



### PART 2: How statistics can be misleading



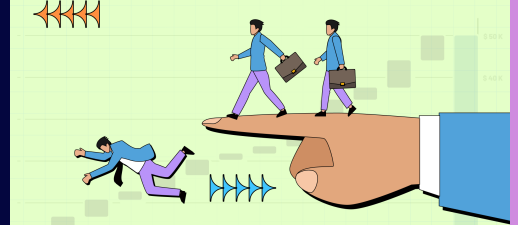
1. Consider the following problem and make a decision according to the data provided. Justify.

Imagine you need to choose between two hospitals for an elderly relative's surgery. Out of each hospital's last 1000 patients, 900 survived at Hospital A, while only 800 survived at Hospital B. Which is the better choice?

2. What could be wrong about one's spontaneous decision in that particular case?

One must take into account the level of health of people arriving at the hospital before making their decision.

## Misleading Statistics in the Media: Examples and How to Spot Them



3. Actually, not all patients arrive at the hospital with the same level of health.

Let's divide each hospital's last 1000 patients into those who arrived in good health and those who arrived in poor health. Hospital A had only 100 patients who arrived in poor health, of which 30 survived. But Hospital B had 400, and they were able to save 210. Which hospital should one choose? Describe your calculations.

Hospital B is the better choice for patients who arrive at hospital in poor health, with a survival rate of 52.5% (210/400). What if your relative's health is good when she arrives at the hospital? Strangely enough, Hospital B is still the better choice, with a survival rate of over 98% ((800 total survivors – 210 survivors originally in poor health)/600 who arrived in good health).

6. What contradiction do the results raise? Do you know what to call such a phenomenon?

How can Hospital A have a better overall survival rate if Hospital B has better survival rates for patients in each of the two groups? Here is a case of Simpson's paradox, where the same set of data can appear to show opposite trends depending on how it is grouped. Or in other words, it is possible to draw two opposite conclusions from the same data depending on how you divide things up. It often occurs when aggregated data hides a conditional variable, sometimes known as a lurking variable, which is a hidden additional factor that significantly influences results. Here the hidden factor is the relative proportion of patients who arrive in good or poor health.

7. What factor should you take into account in the following cases before drawing definitive conclusions? Which is the lurking variable?



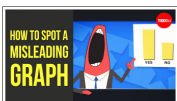
a. One study in the UK appeared to show that smokers had a higher survival rate than non-smokers over a twenty-year time period.

That is until dividing the participants by age group showed that the non-smokers were significantly older on average, and thus more likely to die during the trial period, precisely because they were living longer in general. Here the age groups are the lurking variable and are vital to correctly interpret the data.



b. An analysis of Florida's death penalty cases seemed to reveal no racial disparity in sentencing between black and white defendants convicted of murder.

Dividing the cases by the race of the victim told a different story. In either situation, black defendants were more likely to be sentenced to death. The slightly higher overall sentencing rate for white defendants was due to the fact that cases with white victims were more likely to elicit a death sentence than cases where the victim was black, and most murders occurred between people of the same race.



### PART 3: How to spot a misleading graph

1. In each case, how would you manipulate people into believing what you want? What kind of misleading graph should you resort to? What strategy can you rely upon to distort the information?

Clearly identify the effect expected in order to design the misleading graph.

a. The brand Chevy wants to claim in an ad they make the most reliable trucks in America: 98% of all Chevy trucks sold in the last 10 years are still on the road. But the actual data for the other brands are: 97.5% for Ford, 96.5% for Toyota and 95.5% for Nissan/Datsun. How can they impress customers with such figures?

By distorting the scale, it looks like Chevy trucks are twice as dependable as Toyota trucks. The scale only goes from 95% and 100%; if it went from 0 to 100% the impression would be radically different. This is one of the most common ways graphs misrepresent data: zooming in on a small portion of the y-axis exaggerates a barely detectable difference between the things being compared. And it's especially misleading with bar graphs since we assume the difference in the size of the bars is proportional to the values.

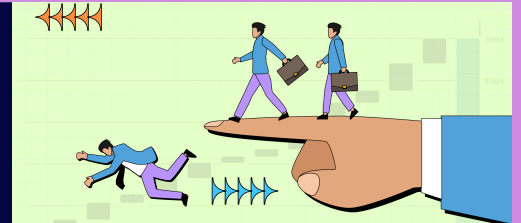
c. To sell more expensive ads, you are required to make it look like the Super Bowl's popularity is exploding by presenting data related to its viewership.

b. Fox News aims at highlighting a dramatic increase in American unemployment from 2008 to 2010.

Here the x-axis is manipulated in two ways. First of all the scale is inconsistent, compressing the 15-month span after March 2009 to look shorter than the preceding six months. Using more consistent data points shows job losses were tapering off by the end of 2009. Moreover the time line starts immediately after the US's financial collapse (subprime mortgage crisis starting in July 2007). These techniques are known as cherry picking. A time range can be carefully chosen to exclude the impact of a major event right outside it. And picking specific data points can hide important changes in between.

d. A global warming sceptic organization wants you to present the evolution of annual global ocean average temperatures from 1880 to 2016.

## Misleading Statistics in the Media: Examples and How to Spot Them



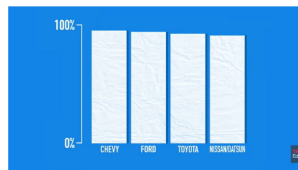
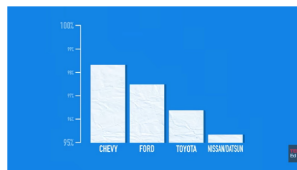
Leaving out relevant data can give a misleading impression. This chart of how many people watch the Super Bowl each year makes it look like the event's popularity is exploding. But it's not accounting for population growth. The ratings have actually held steady because while the number of football fans has increased, their share of overall viewership has not.

A graph can't tell you much if you don't know the full significance of what's being presented. Both of the following graphs use the same ocean temperature data but they seem to give opposite impressions because the first graph plots the average annual ocean temperature from 1880 to 2016, making the change look insignificant. But in fact a rise of even half a degree Celsius can cause massive ecological disruption. This is why the second graph which shows the average temperature variation each year is far more significant.

2. Match the preceding cases a, b, c and d with the most effective misleading strategy or strategies.

**Strategies:** Relevant data left out – Distortion of the y-axis – Selection of specific data points (cherry picking) – Inconsistent scale – Overall context and full significance omitted – Time range carefully chosen (cherry picking)

a. The brand Chevy wants to claim in an ad they make the most reliable trucks in America: 98% of all Chevy trucks sold in the last 10 years are still on the road. But the actual data for the other brands are: 97.5% for Ford, 96.5% for Toyota and 95.5% for Nissan/Datsun. How can they impress customers with such figures?



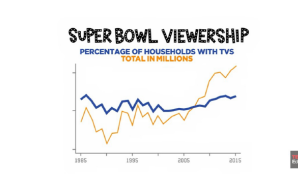
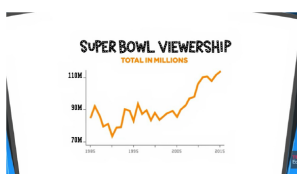
b. Fox News aims at highlighting a dramatic increase in American unemployment from 2008 to 2010.



Strategy : Distortion of the y-axis

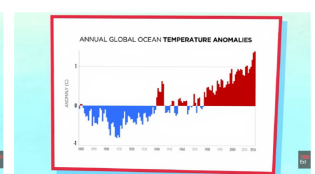
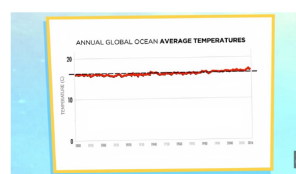
Strategy: Inconsistent scale, Selection of specific data points (cherry picking), Time range carefully chosen (cherry picking)

c. To sell more expensive ads, you are required to make it look like the Super Bowl's popularity is exploding by presenting data related to its viewership.



Strategy : Relevant data left out

d. A global warming sceptic organization wants you to present the evolution of annual global ocean average temperatures from 1880 to 2016.



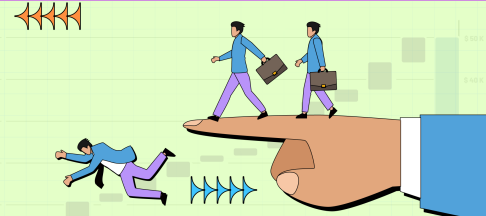
Strategy: Overall context and full significance omitted

3. How would you question the following assertions? What is hidden behind the data?

- Industry trials of drugs are much more likely to show positive results than government-funded studies of the same drugs. Scientists in private companies are more likely to report only their positive trials (cherry picking), and it is then assumed all trials are the same.
- The number of phones sold by a company has dramatically increased. The company resorted to cumulative data (adding each successive input in the data set up > every phone sold ever – the graph will always be rising) rather than annual data (showing the data for each individual year, which could be increasing or decreasing), to make their sales appear larger than they actually are.



# Misleading Statistics in the Media: Examples and How to Spot Them



## Part 4: Can you spot the problems with these headlines?

How is science made more attractive in the news? In medicine notably, there is often a disconnect between news headlines and the scientific research they cover. That's because a headline is designed to catch attention: it is most effective when it makes a big claim.

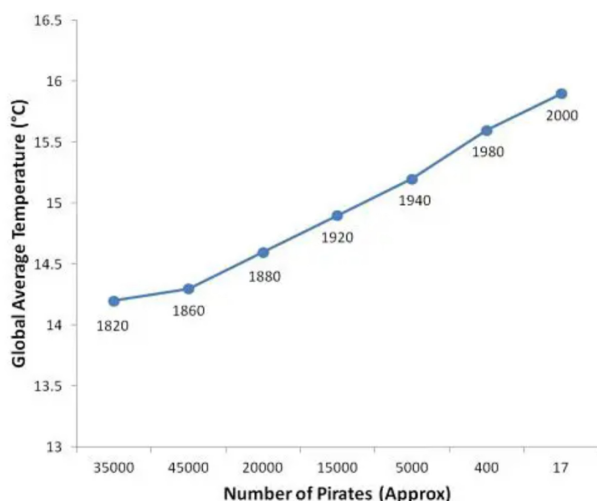
1. Would you be manipulated by the following headlines? Or can you identify a potential distortion of information? What makes these headlines misleading?

	<p>A study finds that the participants taking a vitamin called Healthium had a higher level of healthy cholesterol than those taking a placebo. Their levels became similar to those of people with naturally high levels of this kind of cholesterol. Previous research has shown that people with naturally high levels of healthy cholesterol have lower rates of heart disease.</p>	<p>Answer: The research did not actually investigate whether Healthium reduces heart disease. It only measured Healthium's impact on levels of a particular kind of cholesterol. The fact that people with naturally high levels of that cholesterol have lower risks of heart attacks doesn't mean that the same will be true of people who elevate their cholesterol levels using Healthium.</p>
	<p>This hypothetical study recruits 10 students. Half begin consuming a daily dose of chocolate, while half abstain. As classmates, they all follow the same schedule. By the end of the study, the chocolate eaters are less stressed than their chocolate-free counterparts.</p>	<p>Answer: It's a stretch to draw a conclusion about students in general from a sample of ten. That's because the fewer participants are in a random sample, the less likely it is that the sample will closely represent the target population as a whole (for example, consider the distribution of males and females – the chances of skewing the distribution is higher with a small sample).</p>
	<p>Researchers recruit a group of people who had always skipped breakfast and ask them to start eating breakfast every day. The participants include men and women of a range of ages and backgrounds. Over a year-long period, participants lose an average of five pounds.</p>	<p>Answer: The people in the study started eating breakfast and lost weight – but we don't know that they lost weight because they started eating breakfast; perhaps having their weight tracked inspired them to change their eating habits in other ways. To rule out the possibility that some other factor caused weight loss, we would need a comparison group (who didn't eat breakfast before and continued to skip it).</p>

2. Comment upon the following statement.

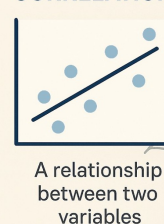
*Pirates are responsible for global warming.*

First one notices that the number of pirates is shrinking over time while the global average temperature is increasing – it is even less likely that plummeting pirates are responsible for global warming than a soaring number of pirates would. But most of all, correlation is not causation.



## Correlation vs. Causation: Untangling the Web of Relationships

### CORRELATION



### CAUSATION

