Data literacy and data manipulation

1. Introduction

The activity What is data? looks at what 'data' is and how it is used by people and businesses to make decisions every single day.

This activity looks in more detail at how that data can be manipulated to present false or misleading information. It doesn't aim to teach numerical skills or statistical methods, but rather to show you how to start thinking about and investigating the context of data.

Data literacy is not just about being able to use numerical skills to manipulate data. It is also about being able to assess and interpret data. Data literacy becomes statistical literacy, or statistical competence, when data is presented in the form of charts, or analysed and communicated in results.

Data is made into statistical representations, such as percentages, charts or graphs, and then meaning is inferred following interpretation of the data. Context, or the story behind the data, can help provide a more complete picture.

2. Misrepresentation of data

Something to bear in mind when looking critically at data is something called misrepresentation, which comes in several different forms:

- **Oversimplification** where a single point is focussed on whilst ignoring other factors, confusing correlation (when two things appear to be related) with causation (when one thing causes another thing to happen), or simply misunderstanding and misrepresenting scientific data.

- **Misleading statistics** this is the misuse of numerical data (intentionally or otherwise) to create a false or misleading message and is commonly used in advertising.
- **False order over random chance** this is where the similarities in a dataset are focussed upon whilst ignoring anything that is too different or contradictory, or when only a small part of a dataset is looked at (for example a few months rather than a whole year, or a small percentage of children in a classroom), leading a false conclusion being presented.

3. Misleading statistics

Data can be analysed and presented in many different ways – such as descriptive statistics (graphical representation of data such as charts and graphs).

The use of statistics to analyse or describe data involves making judgements and inferring meaning. Although data is facts, or given knowledge, these judgements may be open to alternative conclusions which might explain conflicting results and newspaper headlines about similar research.

Numbers are often used in advertising or political campaigns are they are an easy and impactful way of conveying a message. But it is very easy for the numerical data to be misleading. Some very common ways of numerical data being used in misleading ways include:

- numbers taken out of context
- using percentages out of context
- claiming causation instead of correlation

You can read more about these tactics in the article The truth about misleading numbers in advertising.

A famous example of misleading statistics occurred in 2007. The toothpaste manufacturer Colgate was told to remove the claim that 80% of dentists surveyed recommend their product from their advertising as in reality the dentists were allowed to recommend multiple brands, and another brand was recommended almost as much as Colgate. So not exactly a false claim, but certainly a misleading one! (You can read the summary of the report on the BBC News website.)

4. Spotting misleading statistics

It is not always easy to spot when numbers and statistics are being used to mislead or misrepresent issues. But there are some things to think about when presented with statistics such as percentages, even if you are unable to access the data behind them.

1. Are the statistics presented without context? Context gives statistics meaning by situating them within a bigger picture, for example reporting a percentage decrease in the use of fossil fuels in the last year is a good statistic, but how does it sit within the larger historical trend?
2. If they are making comparisons (e.g. compared to last year), are they comparing like with like? For example, is the statistic comparing home energy use in the summer versus the winter (when energy use will always be higher).
3. Does it seem 'too simple'? Is the issue too complicated to be pared down to a simple numerical value?
4. Is the statistic used to try and sell or promote a point of view or product? If so, is it too good to be true?

Is the survey size significant? For example, if a claim is made that 90% of respondents would consider having solar panels installed but only 10 people were surveyed, how much would you trust that statistic?

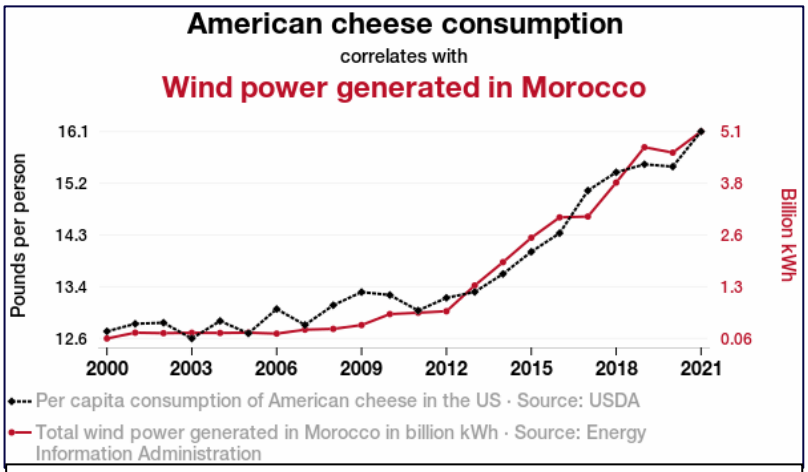
These are just a few of the questions that you can ask when you are presented with a statistic. For more details, the House of Commons Library published the report How to spot spin and inappropriate use of statistics in 2023 which sets out some common examples of statistical spinning and gives some tips on how to spot when this has been done, which can be summarised in the following three prompts:

- Compared to what?
- Since when?
- Says who?

(Bolton and Brooks, 2023)

5. Charts and graphs

Charts and graphs are often used by media outlets as a quick way to visually impart information – but as with other forms of data these can easily be manipulated, either deliberately or by lack of statistical skill.



(Image credit: tylervigen.com/spurious/correlation/2936)

The website Spurious Correlations has a humorous take on this subject, and has created graphs showing how unrelated data can be manipulated to appear to be causally linked. For example, how American cheese consumption has a

correlation with how much wind power has been generated in Morocco over the years.

One of the main things to look at when presented with charts and graphs is the values on the Y-axes (the vertical ones).

- Do the values start at zero, or are they focussed on the 'top' of the data to try and make the differences seem more extreme than they actually are?
- Are the intervals on the Y-axes set at regular intervals or do they double or stretch out the intervals?
- Are the data labels completely missing?

There are some good examples of what misleading charts can look like on the blog post 5 sources of misleading statistics.

6. Data in context

Whilst data itself is usually facts and figures, it has always been collected in a particular way, often for a particular purpose, which gives the data context. This can help with statistical interpretation and also with assessing the data.

After all, if the data is flawed, skewed or taken out of context, then no statistical representation or inference will be reliable or valid. Setting the data in context can help find the story behind it and gives the data meaning.

Looking at data

The following table simply lists categories of countries and their solar energy capacity in the year 2000 and the year 2024, and the difference between the values.

	2000	**2024**	**Change**
High-income countries	1.158	718.232	62030%

Upper-middle-income countries	0.052	1005.743	1942752%
Lower-middle-income countries	0.010	139.866	1438954%
Low-income countries	0.001	1.985	276497%

(IRENA, 2025)

What do you think would help make the data in the table more meaningful? Remember those three prompts from earlier in the activity, as these can help you to spot potentially misleading or incomplete data:

- Compared to what?
- Since when?
- Says who?

Have a think about this and then check our comments.

Querying the data

Looking at the data in the table, there are some things that would help to make it more meaningful, such as:

- What is 'solar energy capacity' and how is it measured?
- Was the capacity measure used in 2000? Was it the same measure in 2024?
- Why was the data collected? What prompted the collection of the data? What might it be used for?
- Does the data include all solar energy producers, or just state owned ones?
- What countries are included in each categorisation and what are the criteria for each definition?

A final, but vital, question to ask is: Who might benefit from this collection and representation of the data? For example, what differences might there be if the data was collected and presented by a 'green group' as opposed to a government agency?

Although data itself may be 'neutral', the collection of data will always be susceptible to some bias due to the decisions made regarding what to count or measure, and the way it is done. But investigating the context of the data may help reveal possible bias.

7. Conclusion

Data literacy is the ability to assess the data and its context. This is important for both consumers of data and those who produce it.

In this activity you have looked at what data literacy means, and you have thought about the questions you can ask when looking at data in order to judge its reliability and validity and to give it context.

If you want to learn more about false information and how to stop its spread, have a look at the other activities in the Spotting and stopping false information pathway.

References

Bolton, P. and Brooks, C. (2023) How to spot spin and inappropriate use of statistics. Available at: https://commonslibrary.parliament.uk/research-briefings/sn04446/ (Accessed: 2 March 2026).

IRENA (2025) 'Total solar capacity' [dataset]. processed by Our World in Data. Available at: https://ourworldindata.org/grapher/installed-solar-pv-capacity?tab=table (Accessed: 27 January 2026.)

Vigen, T. (no date) *Spurious correlation #2,936: American cheese consumption correlates with Wind power generated in Morocco*. Available at: https://tylervigen.com/spurious/correlation/2936_american-cheese-consumption_correlates-with_wind-power-generated-in-morocco (Accessed: 24 April 2026).