

What is data?

1. Introduction

People use data to make decisions every day, whether they realise it or not. If they use the number of votes from customer reviews on a website to decide whether to buy an item or book a certain hotel, then they have made a decision based on data.

But who were those people who voted? How were their votes counted?

This activity will look at what data is, and why it is so important.

Note: The word 'data' is the plural of 'datum'. However, most people tend to use the word 'data' for both the plural and singular form, as we do throughout this activity.

2. What is 'data'?

How would you define 'data' and 'statistics'?

Is there a difference?

As you may have realised, defining these terms is not always clear cut, as the following definitions show.

Data The observations gathered from an experiment, survey or observational study. Often the data are a randomly selected sample from an underlying population. Numerical data are discrete if the underlying population is finite or countably infinite and are continuous if the underlying population forms an interval, finite or infinite. Data are nominal if the observations are not numerical or quantitative, but are descriptive and have no natural order. Data specifying

country of origin, type of vehicle or subject studied, for example, are nominal. (The Concise Oxford Dictionary of Mathematics, 2021)

Datum. Pl. Data A thing given or granted; something known or assumed as fact, and made the basis of reasoning or calculation; an assumption or premise from which inferences are drawn (Oxford English Dictionary, 2022).

Statistics Quantitative data on any subject, usually collected with a purpose such as recording performance of bodies for comparison, or for public record and which can then be analysed by a variety of statistical methods. (The Concise Oxford Dictionary of Mathematics, 2021).

Statistics The practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample (Oxford Dictionary of English, 2015).

3. Using data

Data is a valuable commodity for businesses who can track online trails or purchases and then use this information to their commercial advantage.

Before drawing conclusions based on data, the data itself needs to be looked at critically.

When used correctly, data can be a powerful asset, but it's important to remember that:

- Data by itself is useless: it takes people using the data to give it power and meaning.
- Data is not perfect: Data is often 'incomplete, imperfect, inaccurate or outdated', and should never be taken as a perfect reflection of the world.
- Data needs to be interpreted: Whilst some data requires no prior or specialist knowledge to be interpreted (a bus timetable would be a good

example of this), most data requires further analysis, or some previous knowledge of what is important or not.

- Data doesn't convey power: Just because you know something doesn't mean that you can use that knowledge to make a change.
- You need skills to interpret data: Access to data is easier than ever before, but it is not always easy to know what it means.

Summarised from Gray, 2012

4. Assessing data

Data is frequently presented as numbers – something that has been counted, measured, weighed – but it can also be descriptive (types, names, colours etc.).

Data can come from observations or as the result of experiments.

So, data is facts or knowledge, and statistical methods are used to give meaning to it. This can be through representation or by inferring assumptions about the relationships within the data.

But just as words can be ambiguous, so can numbers. Taking a quotation, or data, out of context can be used to mislead or incorrectly support an argument.

Statistical methods can be used to find patterns in the data and give them meaning so that users do not interpret them incorrectly (or they can also be used to mislead, either purposefully or inadvertently). It is often what is left out that misleads.

5. Conclusion

This activity has broadly defined what 'data' is and outlined how it can be used.

Next steps

If you want to know more about how data can be manipulated to present a misleading or false narrative, have a look at the activity [Data literacy and data manipulation](#).

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

'Data' (2021) in *The Concise Oxford Dictionary of Mathematics*. 6th edn. Available at: <https://www-oxfordreference-com.libezproxy.open.ac.uk/view/10.1093/acref/9780198845355.001.0001/acref-9780198845355-e-736> (Accessed: 28 July 2022).

'Datum. Pl. Data' (2022) in *Oxford English Dictionary*. Available at: <https://www-oed-com.libezproxy.open.ac.uk/view/Entry/47434> (Accessed: 28 July 2022).

Gray, J. (2012) 'What data can and cannot do', *The Guardian*, 31 May. Available at: <https://www.theguardian.com/news/datablog/2012/may/31/data-journalism-focused-critical> (Accessed: 27 January 2026).

'Statistics' (2015) in *Oxford Dictionary of English*. 3rd edn. Available at: https://www-oxfordreference-com.libezproxy.open.ac.uk/view/10.1093/acref/9780199571123.001.0001/m_en_gb0810800 (Accessed: 28 July 2022).

'Statistics' (2021) in *The Concise Oxford Dictionary of Mathematics*. 6th edn. Available at: <https://www-oxfordreference-com.libezproxy.open.ac.uk/view/10.1093/acref/9780198845355.001.0001/acref-9780198845355-e-2696> (Accessed: 28 July 2022).