🌿 Stats in the Wild 🌿

Good Data Guidelines

Finding good quality can be time consuming and yet often represents an essential part of planning statistical activities. 'Good quality' is a flexible term that is related to the purpose of the data, but there are some characteristics that you can look for that will help to ensure that the conclusions you draw from the data can be well-founded.

Here are some things to pay attention to when judging the 'quality' of data:

1. **Reliably sourced** – Is the data published by a reputable organisation? (e.g. known for good statistical practices, affiliated with governments or national organisations, associated with a university or other academic institution)
2. **Transparently collected** – Is there an explanation about how and why the data was collected and by whom (in particular be aware of commercial organisations collecting data for a particular agenda)?
3. **Clearly dated** – Is the time period of the data clearly labelled in such a way that your conclusions can be limited to that particular time period (or extended beyond where there are good reasons for doing so)?
4. **Well-described** – Are variables, units, and data values easy to understand and interpret?
5. **Biases reduced** – Are there are any sources of significant bias (for example in the question posed, in the data collection, in the representations, or in any interpretation) that may undermine conclusions based on the data?
6. **Inclusive** – Where data relates to people or communities, are certain identities or demographics minimised, silenced or erased (this includes the questions posed)?
7. **Ethical** – Was consent given to collect and/or use the data and where relevant was the data sufficiently anonymised?
8. **Relevant** – Is the data useful for the specific issue or problem being explored, without missing out anything important?
9. **Complete** – Is the data free from errors or gaps, or are these clearly labelled as such?

Note: Be aware that data held on the internet can be updated or removed and so where possible download and locally store 'good' data once identified

References

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. Massachusetts Institute of Technology. https://data-feminism.mitpress.mit.edu/

Sebastian-Coleman, L. (2013). *Measuring data quality for ongoing improvement: A data quality assessment framework* (1st edition.). Elsevier. https://www.sciencedirect.com/book/9780123970336/measuring-data-quality-for-ongoing-improvement

Towse, A. S., Ellis, D. A., & Towse, J. N. (2021). Making data meaningful: Guidelines for good quality open data. *The Journal of Social Psychology*, *161*(4), 395–402. https://doi.org/10.1080/00224545.2021.1938811