

Activity ➔ 39

Finding and using datasets

STUDENT HANDOUT

A 'dataset' is a collection of data (gathered by a survey, observation or investigation of a particular phenomenon, for example) that is usually presented in tabular form. 'Big data' refers to extremely large and complex datasets, 'small data' refers to datasets of a manageable volume that are accessible, informative and actionable and 'open data' refers to datasets that are free to use, reuse, build on and redistribute (subject to stated conditions and licence).

Find and use two datasets that are relevant to your dissertation. These could be datasets that help with your background research or datasets that help to explain emerging results, for example. As you use the datasets, answer the following questions. We will use these questions as a basis for discussion when we next meet.

1. Where did you find the datasets?
2. How easy were they to find?
3. Did you encounter any difficulties when searching for relevant datasets?
4. How well were the data presented?
5. Were they easy to understand?
6. What method(s) did you use to explore the data?
7. Did you encounter any difficulties when exploring the data?
8. What tasks did you perform (sort, filter, code, query or annotate, for example)?
9. How successful were these tasks?
10. Did you encounter any problems when performing these tasks?
11. How useful did you find the information? Is the information relevant to your research?
12. Did you find any datasets that weren't suitable or were inappropriate? If so, why was this?
13. What tips and advice would you give to peers who are thinking about finding and using datasets?

Learning outcome: By the end of this activity you will understand how to find, use and critique datasets that are relevant to your assignments, projects and research.