Social Research Methods

Extension Activity: Introduction to the UK Data Archive

Secondary data analysis involves using data that already exist. There are a wealth of social science data resources available for researchers to utilise for their own novel research projects. The UK is often described as 'data rich but analyst poor' as we have a large volume of existing high quality data resources, but we lack researchers with the skills or interest to exploit these resources.

Existing data resources include high quality, nationally representative social survey data collections which cost millions of pounds to collect and curate. These data are of far higher quality than a single researcher or small research team would be able to collect by themselves.

This activity introduces you to how to access existing data resources which could be used in your own research (e.g. your dissertation).

The UK Data Archive

The UK Data Archive is an internationally acknowledged centre of expertise in acquiring, curating and providing access to social science and humanities data (both quantitative and qualitative).

The UK Data Archive was founded in 1967, at the University of Essex, with the support of the then Social Science Research Council, with the aim of curating high-quality research data for analysis and reuse. The Data Archive is now a significant part of their UK data infrastructure.

The UK Data Archive can be found here: http://data-archive.ac.uk/home.

- Please familiarise yourself with the UK Data Archive website and read about the services it offers.

Data Holdings

Data Catalogue

There are thousands of data resources available for researchers to access at the UK Data Archive. Take a moment to look through the resources which are available. Take care to read the details of the data to understand whether they are qualitative or quantitative in nature.

Information on key data sets is available here:

https://www.ukdataservice.ac.uk/get-data/key-data

Information on recently deposited data sets in available here:

https://tinyurl.com/ybbvyfv7

You can manually search the archive here:

https://tinyurl.com/ybbvyfv7

- Try searching for data which matches your personal research interests.

Data Documentation

The UK Data Archive has extensive data documentation, carefully reading this documentation will allow you to understand what variables are in the data and how the data are collected. Before you use a piece of data you should carefully read this documentation in order to understand how the data have been produced.

Let's take a look at the documentation for the Health Survey for England (2016). Visit the main catalogue page here:

https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8334.

- Click on the 'Documentation' tab.

- Under the 'Documentation' tab, view the 'HSE User Guide'. Take a look at this document and find the details of how the data were selected (e.g. how were respondents selected to take part in this survey).

- Under the 'Documentation' tab, view the 'HSE Questionnaires and Showcards'. Can you find the questions the respondents were asked about their general wellbeing?

If you were to use an existing data resource in your own research you would want to report details of how the data were collected (e.g. sampling and survey mode), and cite the data documentation in the same way you would cite a journal article or book.

Study Numbers

One of the easiest ways to find data is to search using the UK Data Archive Study Number.

If you look at the Health Survey for England (2016) catalogue page again you will see the study number listed (SN8334):

https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8334#!/details

When you are working with a data set it is useful to note this number, you can then simply search for the study number and you will return to the required catalogue page.

- Try it, search for SN8334: https://tinyurl.com/y86zscjg.

Citing Data

When using existing data, researchers should acknowledge and cite the data in the same way you would cite a journal article or book.

Data citation:

- Acknowledges the work of the data collectors,
- Makes identifying the data you have used easier,
- Promotes the reproducibility of research results (i.e. other people can repeat your research),
- Allows the use of data to be tracked (i.e. data funders can track how their data is being used).

Helpfully, the UK Data Archive tells you how to cite the data resources it provides. Revisit the Health Survey for England (2016) catalogue page. Scroll down and you will see the citation for this study, you can also change the format to your preferred style.

In your text you would cite the study using the first institution listed as the 'author' and the date given. For example:

In this research project I used the 2016 Health Survey for England (University College London, 2018).

Quiz: Use the links above to search the data catalogue and answer the questions below (answers are at the end of this document).

1) You are interested in undertaking a piece of longitudinal research which follows the lives of children born in the year 1970. You would like to link parental social class at the child's birth to the social class the child attains in adulthood. Identify a potential data set.

2) You are interested in gender and attitudes towards abortion in the UK. Identify a potential data set.

3) You are interested in age and the fear of crime in England and Wales. Identify a potential data set.

4) You are interested in ethnicity and gender differences in the time spent on housework in the UK. Identify a potential data set.

5) What is the study number of the National Survey of Sexual Attitudes and Lifestyles (1990)?

6) How would you cite the Learning at Work Survey (2004) in your reference list in Vancouver format?

Data Access

Bone fide researchers, and students such as yourselves, are able to access many data resources from the UK Data Archive. These are real data about real people and should be treated ethically and with respect. You should carefully read any terms and conditions you agree to when downloading data. You should never share any data you download.

Please read the UK Data Archive data access policy here:

https://www.ukdataservice.ac.uk/get-data/data-access-policy/legal-definitions

All users are required to agree to the terms and conditions pertaining to the use of the data they download. Read the terms of the UKDA End User License here:

https://www.ukdataservice.ac.uk/get-data/how-to-access/conditions

Some data resources cannot simply be downloaded, for some more secure data resources you would need to access them via a secure lab, or undergo additional training and security checks before access is granted.

Take a moment to register with the UK Data Archive. Register using your university credentials (i.e. your email address which ends in '@york.ac.uk'). If you do not do this your registration will not work.

- Visit the main UK Data Service Webpage (https://www.ukdataservice.ac.uk/) and click on 'Register' at the top right hand side of the page. Follow the instructions to complete your registration.

Note: You are from a UK institution of higher or further education so please select this option when prompted.

Secondary Data Analysis Practice

Please select one of the data sets listed below. Download your chosen data set in SPSS format (under the 'access' tab, do not click on 'access online', click on 'Download' next to SPSS). Take a moment to explore your chosen data set and its documentation.

- SN 6891: British Crime Survey 2007-2008: Unrestricted Access Teaching Dataset.
- SN 7613: Census 2011 Microdata Teaching File for England and Wales.
- SN 7913: Opinions and Lifestyle Survey, Well-Being Module, April-May 2015: Unrestricted Access Teaching Dataset.

- Think of potential research questions you could answer using this data set.

Complete the following tasks:

- Produce some descriptive statistics for a scale variable (e.g. calculate mean, median, mode, maximum, minimum);
- Produce a graph of a scale variable;
- Produce some descriptive statistics for a categorical variable (e.g. describe the number of people in each category);
- Produce a graph of a categorical variable;
- Recode a scale variable into categories;
- Recode a categorical variable into fewer categories;
- Produce a contingency table to explore the association between two categorical variables;
- Produce a graph of your contingency table.

Data Holding Quiz Answers

Note: There are potentially multiple data sets which could be used to answer these questions. The answers below identify suitable data sets, but you may have identified additional suitable resources.

1) The most plausible data resource for this enquiry will be the [1970 British Cohort Study](#).

2) The most plausible data resource for this enquiry will be the [British Social Attitudes Survey](#).

3) The most plausible data resource for this enquiry will be the [Crime Survey for England and Wales](#).

4) The most plausible data resource for this enquiry will be the [United Kingdom Time Use Survey](#).

5) SN3434

6) Fuller, A., Felstead, A., Unwin, L. Learning at Work Survey, 2004. [data collection]. UK Data Service, 2008 [Accessed 1 October 2018]. Available from: http://doi.org/10.5255/UKDA-SN-6084-1