

Understanding the Enumerated World

Making Sense of Data as an Information Source

Alexandra Cooper, Data Services Coordinator, Queen's University, coopera@queensu.ca; Elizabeth Hill, Data Librarian, Western University, ethill@uwo.ca; Kristi Thompson, Research Data Management Librarian, Western University, kthom67@uwo.ca

NUTRITION INFORMATION

This recipe is a guide to preparing an instructional session aimed at postsecondary students in the social or health sciences or related disciplines on locating, evaluating, and using secondary data sources as information resources.

Who collects data? Where can you access them? Why are data available on some topics and not others? Why are some statistics available at a detailed level of geography and others only nationally? What are some key limitations of official statistics, and where can information be found to fill in the gaps? This session uses these questions to encourage students to consider how data are used as information sources.

Governments are a primary creator and distributor of statistical information. They collect data to help develop policy and to allow for planning and assessment of services. Thinking about these purposes can help us understand why some topics are covered in depth and others are covered in less detail or not at all. We will draw distinctions between administrative data and survey data, with examples of how both types of data play key roles in generating government statistics. Topics include privacy and confidentiality,

data suppression, and the role of sampling with regard to survey data. This session will explore a range of data dissemination formats from reports and summary tables to full data sets.

TARGET AUDIENCE AND NUMBER SERVED

Up to 40 students in a classroom setting. This session has been developed to be palatable to data novices. It best serves students in social science, business, or health science courses, which frequently use data as a source of information: to profile a community, write a business plan, find prevalence of a health condition, determine audience for a program, or investigate a social problem.

LEARNING OUTCOMES

Students will

- describe what data are and their importance in society and research
- identify government and other organizations as sources of secondary data
- evaluate usefulness of secondary data sources for research

COOKING TIME

- *Preparation*: one day.
- *Lecture and workshop*: one class session

that is 1½ to 2 hours in duration. Lecture should take no more than one hour. Remaining time should be used to introduce and start the assignment.

- *Assignment*: Assignment should be planned with the faculty member in advance of the session. Students should be given an appropriate length of time outside of class to complete the assignment, working individually or in small groups, depending on the size of the class. The librarian should be available for consultation with the students by appointment over this period.

DIETARY GUIDELINES

This session fits into ACRL's *Framework for Information Literacy for Higher Education*. The focus is specifically tied to the second frame, Information Creation as a Process: "Information in any format is produced to convey a message and is shared via a selected delivery method. The iterative processes of researching, creating, revising, and disseminating information vary, and the resulting product reflects these differences."

The goal of this session is to help students develop into informed and critical consumers of information by teaching them how to locate, access, and evaluate data.

INGREDIENTS***Kitchen Equipment***

Computer with an internet connection. If each student has their own computer, they can follow along more easily. Consider taking screenshots of interactive sites in case there are internet issues.

Ingredients

Data resources to highlight based on the subject being taught, the content of the course, and relevant geography. Students are generally more engaged when the instructor shows them data on their hometown.

PREPARATION

The lecture portion highlights examples of the three levels of data collection: international, national, and local. Determine the resources to highlight in consultation with the faculty member leading the course. See the Additional Resources section for a list of suggested sites of data sources.

A key ingredient is the national census (American, Canadian, United Kingdom, or whichever is relevant). You will use part of the lesson to briefly demonstrate how to search, browse, and download data from two or three sites. We suggest selecting the census, an international data source such as the World Development Indicators at the World Bank, and possibly an additional site that is relevant to the course.

The assignment should be developed in conjunction with the faculty member and should

be directly relevant to course outcomes. One option that works well is to have students profile a community, group, area, or country on a topic of interest, drawing information from at least two of the three levels of data collection (international, national, local) that will be identified.

INSTRUCTIONS***Introduction: What Are Data?***

The introductory part of the lesson is conducted as a traditional lecture with slides, with frequent stops to engage the class in discussion. Start with a simple definition of *data* to pique students' interest and introduce the topic in a humorous and nonintimidating way. Start with the question "What are data?" Encourage students to offer examples.

After some discussion, explain: Numeric data come from people counting things, and most of the data we deal with in the social sciences are either administrative data or survey data. Differentiate the following concepts:

- Administrative data are data that are collected as by-products of administering something (think hospital admissions or tax records).
- Survey data are data that are collected specifically for information or research purposes, by formally asking questions of a population (or sample) of respondents.
- A census is a survey that tries to collect information about every member of some population.

Data are released as microdata or macrodata. Explain the difference.

You may choose to highlight microdata examples for a course with an analysis component, or primarily emphasize macrodata released in tables and reports for courses where the student is not expected to conduct an analysis. Many of the sites we suggest feature both. (See Additional Resources).

Section: Where or Who Do Data Come From?

Explain the differences between public and private data sources:

- Public sources such as municipal, state, and national governments; governmental organizations like the UN and the World Bank. Nonprofit organizations and academic researchers are other sources of public data.
- Private sources such as companies, although these data are generally very expensive or proprietary and not released. Some limited data may be made available in shareholder reports and similar documents.

Publicly funded institutions have a mandate to spend their money toward certain goals and are held accountable to the public. Private institutions and businesses have no such mandate and are not accountable beyond what is required by law. Therefore, most publicly available data come from public institutions.

Instruct the class to begin a data search with these questions:

- Who cares about this topic?
- Who has access to this population?

- Who has funding to collect these data and a mandate to release the data?

Encourage them to think of service providers and advocacy organizations in addition to governmental sources.

Highlight that data on many topics may not be available for reasons including

- Organizations interested in the topic did not have the funding to collect it.
- Data were collected but not released for privacy or ethical reasons.
- Population is difficult to collect data on (e.g., the homeless population).

Use examples to demonstrate techniques like extrapolating from a similar population or a different level of geography when a data search is not producing results.

Section: Surveys, Samples, and Suppression

Explain that data collection methodology influences the level of detail that will be available in the statistics derived from those data. Sample surveys are designed to gather information on particular populations and subpopulations, and geographic precision and other detailed information about individuals is often suppressed or generalized in microdata to preserve respondents' confidentiality. A larger survey will give more reliable estimates for smaller subpopulations. Surveys are often designed to obtain accurate estimates for specific subgroups and geographic regions. National health surveys provide a good example of this.

Section: Government Data

Learning about why data are collected will help students better understand the data they are using and develop strategies to find them. Explain that governments collect data to allow for planning and assessment of policies and services. Specific data collection programs survey as few people as possible to collect enough data for specific government needs. Governments also collect administrative data from sources such as tax forms and birth or death records.

Highlight examples such as a health needs assessment survey. In Canada, community health needs are managed by regional health authorities, so the Canadian Community Health Survey collects and releases data at this geographic level. The US National Health Interview Survey similarly collects and releases data for the four census regions (Northeast, Midwest, South, and West), and within census regions by areas determined by metropolitan and nonmetropolitan status.

Section: Top Down, Bottom Up

Differentiate between the following techniques for locating data on a small geographic area.

- *Top down:* Look at national data and drill down to a local area. Demonstrate searching census data for information on a state or city.
- *Bottom up:* Start with locally collected data, for example from a municipality or a local advocacy organization. Provide locally relevant examples to illustrate the point.

REVIEWS/ASSESSMENT STRATEGY

The session works best if it is tied directly to a research assignment where the student needs to find statistical information; the students' success in doing so can serve as a proxy for evaluating the impact of the session.

CHEF'S NOTES

These types of sessions have been a good fit across multiple disciplines. They can easily be adapted by selecting appropriate data sources; for example, by focusing on sources of environmental, economic, or health data.

ADDITIONAL RESOURCES

Suggested sites for secondary data

- International
 - European Union, Eurostat, <https://ec.europa.eu/eurostat/>
 - Organisation for Economic Co-operation and Development, OECD Data, <https://data.oecd.org>
 - United Nations, UNdata, <https://data.un.org>
 - World Bank Open Data, <https://data.worldbank.org>
- National
 - Statistics Canada, <https://www.statcan.gc.ca/eng/start>
 - UK, Office for National Statistics, <https://www.ons.gov.uk>
 - US, Data.gov, <https://www.data.gov>
- Census
 - Statistics Canada, Census of Population <https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>
 - UK, Office for National Statistics, Census, <https://www.ons.gov.uk/census>

- US Census Bureau, <https://www.census.gov>
- Local
 - United States
 - Chicago Data Portal, <https://data.cityofchicago.org>
 - New York City Open Data, <https://opendata.cityofnewyork.us/data/>
 - Open Government, Data.gov, <https://www.data.gov/open-gov/>
 - Lists of US city and county open data sites
 - Canada
 - Toronto Open Data, <https://open.toronto.ca>
 - Vancouver Open Data Portal, <https://opendata.vancouver.ca/pages/home/>