

# UC San Diego

## Books and Book Chapters

### **Title**

Three-Step Data Searching

### **Permalink**

<https://escholarship.org/uc/item/7b52z0c2>

### **ISBN**

978-0-8389-3925-3

### **Author**

Sklar, Annelise

### **Publication Date**

2022

### **Supplemental Material**

<https://escholarship.org/uc/item/7b52z0c2#supplemental>

**THE**  
**DATA LITERACY**  
**COOKBOOK**

edited by Kelly Getz and Meryl Brodsky



# ***The Data Literacy Cookbook***

*edited by Kelly Getz and Meryl Brodsky*

Association of College and Research Libraries

*A division of the American Library Association*

Chicago • 2022

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

The ACRL Cookbook series was conceived of and designed by Ryan Sittler and Doug Cook.

Other books in this series:

*The Scholarly Communications Cookbook*, edited by Brianna Buljung and Emily Bongiovanni

*The Teaching with Primary Sources Cookbook*, edited by Julie M. Porterfield

*The Library Outreach Cookbook*, edited by Ryan L. Sittler and Terra J. Rogerson

*The Critical Thinking about Sources Cookbook*, edited by Sarah E. Morris

*The Sustainable Library's Cookbook*, edited by Raymond Pun and Gary L. Shaffer

*The Library Assessment Cookbook*, edited by Aaron W. Dobbs

*The First-Year Experience Cookbook*, edited by Raymond Pun and Meggan Houlihan

*The Discovery Tool Cookbook*, edited by Nancy Fawley and Nikki Krysak

*The Embedded Librarian's Cookbook*, edited by Kaijsa Calkins and Cassandra Kvenild

Library of Congress Control Number: (TO COME)

©2022 by the Association of College and Research Libraries, a division of the American Library Association. All rights reserved except those which may be granted by Sections 107 and 108 of the Copyright Revision Act of 1976. Printed in the United States of America.

26 25 24 23 22 5 4 3 2 1

# TABLE OF CONTENTS

vii Introduction

## SECTION 1. INTERPRETING POLLS AND SURVEYS

3 **[[Ch1]]Survey Literacy: A Skills-Based Approach to Teaching Survey Research**  
*Jesse Klein*

6 **[[Ch2]]Setting the Scene with Surveys: Using Polling Software to Demonstrate Primary and Secondary Data**  
*Wendy G. Pothier*

9 **[[Ch3]]The Mini-study: A Three-Part Assignment for Original Data Creation, Summation, and Visualization**  
*William Cuthbertson, Lyda Fontes McCartin, and Sara O'Donnell*

## SECTION 2. FINDING AND EVALUATING DATA

17 **[[Ch4]]Three-Step Data Searching**  
*Annelise Sklar*

21 **[[Ch5]]Transforming Research Questions into Variables: A Recipe for Finding Secondary Data**  
*Alicia Kubas and Jenny McBurney*

25 **[[Ch6]]Sweeten the Search: Discover Data for Reuse with a Tool That Links Publications to the Underlying Data**  
*Elizabeth Moss*

30 **[[Ch7]]The Most Vital Statistics: Finding and Analyzing Historical Mortality Rates**  
*Alisa Beth Rod and Jennie Correia*

34 **[[Ch8]]Understanding the Enumerated World: Making Sense of Data as an Information Source**  
*Alexandra Cooper, Elizabeth Hill, and Kristi Thompson*

38 **[[Ch9]]Looking at Data**  
*Kay K. Bjornen*

43 **[[Ch10]]Interrogating the Data: What Data Sets Can and Cannot Tell Us**  
*Kristin Fontichiaro*

46 **[[Ch11]]Data Zines: A Hands-On Approach to Community Curiosities**  
*Tess Wilson*

49 **[[Ch12]]On the Hunt: Understanding and Analyzing GSS Data Extraction for Incorporation within Sociological Research Projects**  
*Amy Dye-Reeves*

52 **[[Ch13]]Using Statistics to Define the Problem: Data and Service Learning**  
*Amy Harris Houk and Jenny Dale*

56 **[[Ch14]]Data and Statistics in the News and Media**  
*Kaetlyn Phillips*

## SECTION 3. DATA MANIPULATION AND TRANSFORMATION

61 **[[Ch15]]A Kinesthetic Approach to Data: Moving to Understand Nominal, Ordinal, Interval, and Ratio Relationship in Data**  
*Wendy Stephens*

64 **[[Ch16]]Text Mining Charcuterie Board**  
*Yun Dai and Fan Luo*

67 **[[Ch17]]Anyone Can Cook (R)! Open Data with R, a Five-Week Mini-mester**  
*Jay Forrest and Ameet Doshi*

70 **[[Ch18]]Software Carpentry AI Dente: Rendering Tech Training for Online Artisans**  
*Peace Ossom-Williamson, Shiloh Williams, and Hammad Rauf Khan*

73 **[[Ch19]]A Recipe for Improving Online Instruction for the Carpentries**  
*Kay K. Bjornen and Clarke Iakovakis*

## SECTION 4. DATA VISUALIZATION

79 **[[Ch20]]Correlation Does Not Equal Causality: Introducing Data Literacy through Infographics and Statistics in the Media**  
*Nick Ruhs*

83 **[[Ch21]]Pies, Bars, Charts, and Graphs, Oh My! A Data Visualization Appetizer**  
*Haley L. Lott*

86 **[[Ch22]]Data Visualizations: The Good, the Bad, and the Ugly**  
*Kaetlyn Phillips*

89 **[[Ch23]]Seasonal Visual Literacy: Using Current Events to Teach Data and Spatial Literacy Skills with Adaptable LibGuides**  
*Jacqueline Fleming and Theresa Quill*

93 **[[Ch24]]To Visualize Is to Experience Data**  
*Chelsea H. Barrett and Gerard Shea*

97 **[[Ch25]]Upping the Baseline for Data Literacy Instruction**  
*Jessica Vanderhoff*

## Table of Contents

101	<b>[[Ch26]]A Literacy-Based Approach to Learning Visualization with R's ggplot2 Package</b> <i>Angela M. Zoss</i>	136	<b>[[Ch36]]Seven Weeks, Seven DMPs: Iterative Learning around Data Management Plan Creation</b> <i>Emma Slayton and Hannah C. Gunderman</i>	179	<b>[[Ch46]]Data Decisions and the Research Process in the Sciences and Social Sciences</b> <i>Nicole Helregel</i>	
104	<b>[[Ch27]]Build Your Own Data Viz Pizza: A Modular Approach to Data Visualization Instruction</b> <i>Rachel Starry</i>	140	<b>[[Ch37]]Equitable from the Beginning: Incorporating Critical Data Perspectives into Your Research Design</b> <i>Jodi Coalter, David Durden, and Leigh Amadi Dunewood</i>	182	<b>[[Ch47]]Financial Data for Economics Students</b> <i>Jennifer Yao Weinraub</i>	
108	<b>[[Ch28]]Veggie Pizza: Choosing a Data Visualization Tool</b> <i>Rachel Starry</i>	<b>SECTION 6. GEOSPATIAL DATA</b>			183	<b>[[Ch48]]Stuffed Shiny App with Business Intelligence</b> <i>Yun Dai and Fan Luo</i>
111	<b>[[Ch29]]Four-Cheese Pizza: Color and Accessible Design</b> <i>Rachel Starry</i>	147	<b>[[Ch38]]Challenge Accepted: Introducing Geospatial Data Literacy through an Online Learning Path</b> <i>Joshua Sadvari and Katie Phillips</i>	189	<b>[[Ch49]]Fast Casual Marketing Strategies</b> <i>Juliann Couture, Halley Todd, and Natalia Tingle Dolan</i>	
114	<b>[[Ch30]]Data Visualization using Web Apps in a Rainbow Layer Cake</b> <i>Yun Dai and Fan Luo</i>	151	<b>[[Ch39]]GIS for Success Series: Learning the Basics of QGIS Workshop</b> <i>Kelly Grove</i>	192	<b>[[Ch50]]When and Where: A Framework for Finding and Evaluating Social Science Data for Reuse</b> <i>Ari Gofman</i>	
117	<b>[[Ch31]]Graphical Abstracts: Creating Appetizing Infographics for Your Research Article</b> <i>Aleshia Huber</i>	154	<b>[[Ch40]]GIS for Success Series: Let's Make a Map in QGIS Workshop</b> <i>Kelly Grove</i>	197	<b>[[Ch51]]Data Literacy Layered Lasagna for Preservice Teachers</b> <i>Brad Dennis and Allison Hart-Young</i>	
<b>SECTION 5. DATA MANAGEMENT AND SHARING</b>		157	<b>[[Ch41]]Statistical and Geospatial Literacy for Integrative Genetics</b> <i>Jay Forrest and Chrissy Spencer</i>	<b>SECTION 8. DATA LITERACY OUTREACH AND ENGAGEMENT</b>		
123	<b>[[Ch32]]Making File Names for Digital Exhibits</b> <i>Kate Thornhill and Gabriele Hayden</i>	161	<b>[[Ch42]]Web Map Layer Cake: Teaching Web Mapping Skills with Leaflet for R</b> <i>Sarah Zhang and Julie Jones</i>	203	<b>[[Ch52]]Data Visualization Day: Promoting Data Literacy with Campus Partners</b> <i>Wenli Gao</i>	
126	<b>[[Ch33]]Data Management Failures: Teaching the Importance of DMPs through Cautionary Examples</b> <i>Richard M. Mikulski</i>	<b>SECTION 7. DATA IN THE DISCIPLINES</b>			206	<b>[[Ch53]]Getting Messy Ourselves: An Experiential Learning Curriculum for Subject Librarians to Engage with Data Literacy</b> <i>Adrienne Canino</i>
131	<b>[[Ch34]]Low-Fat Research Data Management</b> <i>Elizabeth Blackwood</i>	167	<b>[[Ch43]]Data in Context: How Data Fit into the Scholarly Conversation</b> <i>Theresa Burress</i>	211	<b>[[Ch54]]Research Data Management Stone Soup: Gauging Team Competencies</b> <i>Michelle Armstrong, Megan Davis, Ellie Dworak, Yitzhak "Yitzy" Paul, and Elisabeth Shook</i>	
134	<b>[[Ch35]]Managing Qualitative Social Science Data: An Open, Self-Guided Course</b> <i>Sebastian Karcher and Diana Kapiszewski</i>	171	<b>[[Ch44]]Let the Dough Rise! Integrating Library Instruction in a Digital Humanities Course</b> <i>René Duplain and Chantal Ripp</i>			
		175	<b>[[Ch45]]Ethics and Biodiversity Data</b> <i>Rebecca Hill Renirie</i>			

- 214 **[[Ch55]]Data Literacy Family Style: Full-Day Professional Development**  
*Molly Ledermann, Emilia Marcyk, Terence O'Neill, and Dianna E. Sachs*
- 217 **[[Ch56]]Everyone Is Welcome at the Table: Outreach for Data Management and Data Literacy in Research Assignment Design**  
*Shannon Sheridan and Hilary Baribeau*
- 220 **[[Ch57]]Seasoning and Simmering: Cultivating Data Literacy Skills through an Open Data Hackathon**  
*Peace Ossom-Williamson*
- 223 **[[Ch58]]From Soup to Nuts: Finding Your Way around the Data Services Buffet**  
*Jane Fry and Chantal Ripp*
- 226 **[[Ch59]]Teaching Data Literacy and Computational Thinking in Educational Technology**  
*Lesley S. J. Farmer*
- 247 **[[Ch64]]Data as Curation: Framing Data Creation as a Critical Practice through Collections-Based Research Inquiry**  
*Gesina A. Phillips, Tyrica Terry Kapral, Matthew J. Lavin, and Aaron Brenner*
- 253 **[[Ch65]]Quantitative Data Skills for Undergraduates: A Seminar Series for Social Science Students**  
*Whitney Kramer and Amelia Kallaher*

**SECTION 9. DATA LITERACY PROGRAMS AND CURRICULA**

- 231 **[[Ch60]]Cooking Up a Data Literacy Course**  
*Claire Nickerson*
- 238 **[[Ch61]]Baking a Data Layer Cake: Scaffolding Data Skills through Video Vignettes**  
*Shannon Sheridan*
- 241 **[[Ch62]]Building Data Literacy through Scaffolded Workshops: Experiences and Challenges**  
*Jiebei Luo and Yaqing (Allison) Xu*
- 244 **[[Ch63]]Data Literacy Appetizers: LibGuide Data Instruction Modules for Undergraduates**  
*Beth Hillemann and Aaron Albertson*

## Three-Step Data Searching

Annelise Sklar, Assistant Director, Scholarship Tools and Methods Program, University of California San Diego Library, [asklar@ucsd.edu](mailto:asklar@ucsd.edu)

### NUTRITION INFORMATION

This lesson is designed for upper division undergraduate students or early stage graduate students who are required to find and analyze social sciences data for original research, such as for an honors or capstone project. It is meant to follow and build on a literature searching instruction session. The purpose of this lesson is to introduce students to the concepts they need to find reusable data. After a brief introductory lecture, the students will conduct a “reference interview” with themselves to identify avenues for finding research data.

### TARGET AUDIENCE AND NUMBER SERVED

This lesson is most interactive in discussion-sized courses, but it can be scaled up to a lecture.

### LEARNING OUTCOMES

By the end of the session, students will be able to

- explain how data are collected and created by different types of researchers or organizations and described in different types of publications
- identify and use search tools to find data based on discipline, format, or creator’s affiliated organization type

### COOKING TIME

This lesson is meant to follow and build on a library instruction session on literature searching. Walking through the points on this worksheet with students requires about 15–30 minutes. More time can be added for students to apply the questions to their own research topics, for a librarian to demonstrate databases, or for students to explore recommended databases. For the whole session, 45 minutes to an hour is generally recommended.

### DIETARY GUIDELINES

This lesson touches on the following frames from ACRL’s *Framework for Information Literacy for Higher Education*:

- *Information Creation as a Process*: Students will be prompted to think about why data are collected and by whom.
- *Information Has Value*: The instructor will introduce the concepts of open data versus licensed data and explain that students simply may not have access to some data due to restricted use agreements and cost-prohibitive products.
- *Research as Inquiry*: The instructor will reiterate the importance of the literature review and building on existing disciplinary knowledge. Students will be reminded to review publications’ methodology and data sections for references to data that have been used in previous

studies as well as known limitations of available data.

- *Searching as Strategic Exploration*: Students will understand that data can be described and published in a variety of sources and that they will likely have to search multiple resources using a variety of search terms and strategies.

### INGREDIENTS

- Handout/worksheet with definitions of key vocabulary relating to data and questions for students to consider (see appendix)
- Examples of library-licensed or open data sets that match student topics; for example, ICPSR, the Dataverse Project, DataPlanet, Social Explorer, [Data.census.gov](http://Data.census.gov), ProQuest Statistical Insight, and so on.
- Presentation slides or other visual aid with concepts from handout (optional)
- Computers with internet connections for students to practice searching individually or in small groups (optional)

### PREPARATION

This recipe works best when supplemented with a few targeted examples that demonstrate different types of data and sources. For example, political science students who are comparing countries might be introduced to panel data from international organizations



such as the World Bank or OECD, the Correlates of War data set, or the Europa World Plus database. Urban studies students researching their local community might be taught about census data, local government data, or databases such as Social Explorer and SimplyAnalytics. Ideally, the librarian will request and receive students' topics ahead of time in order to choose sources that match individual students' research questions.

The librarian should customize the handout and slides to promote available data sources that match course topics.

### INSTRUCTIONS

Pass out the handout to students at the beginning of class so that they may follow along.

Begin by sharing brief definitions of the word *data*; descriptions of different types and formats; a differentiation between microdata, aggregate data, and statistics; and explanations of the terms *data set* and *repository*.

Following the handout and lecture:

1. Prompt students to think about the kinds of data that would be useful to their projects and to jot down ideas about who might have collected or created these data (e.g., government agencies, nonprofit organizations, private business or industry, or academics). This part is often tricky for novice researchers, so instructors may want to phrase the question in such a way that the students consider each potential creator one by one.

- » For example, the instructor may ask, "Based on what you know, do you think the government would have collected these data?" Depending on the nature of the data, the instructor may follow up by suggesting students think about government forms they've filled out and whether their question was asked, such as whether their voter registration card asked about ethnicity.
  - » After a pause, the instructor may follow with, "Is there a nonprofit [or nongovernmental] organization that might be tracking these data? This might be particularly likely for data that are too controversial for a government to collect—or be trusted to collect impartially."
  - » Then, "Is this information a business might want to collect, or an industry group on behalf of businesses, maybe for marketing or monitoring the competition? Or are these data someone might collect and sell, like a professional public opinion poll?"
  - » And finally, "Is this something academic researchers might collect, maybe in a survey or experiment?"
2. Explain that different kinds of researchers publish their results in different types of publications (scholarly articles, reports, etc.) and recommend students pay attention to how data are described or cited in these publications. Searching literature is one way of finding data. If students have already found relevant articles, prompt them to look at the methodology section to see what they can learn about the data.

Connecting to the first question, point out that academics typically publish in scholarly articles and books, but experts who work for the government, think tanks, nonprofits or nongovernmental organizations, and intergovernmental organizations often communicate through reports and websites published by their organizations (this is called *gray literature*). Likewise, industry and commercial research may be distributed through trade publications. No matter the publication type, students should find out where the data originated. They should try to determine whether the authors created their own data via original research, or whether they are using data from another source and there is a data citation, or whether the article merely gives the names of the individuals or organizations that collected the data.

3. Finally, explain that, depending on their discipline or even their institutional culture, creators of data have different distribution norms. Data may be available in institutional repositories, in licensed databases, or on an organization's or researcher's web page. There also may be accessibility differences. The data may be available only upon request, only for a fee (possibly with a restrictive license), or sometimes not at all, depending on privacy issues, proprietary information, or a researcher's individual inclination or grant specifications.

If time permits and students have developed their topics, the librarian may demonstrate

select databases and search tools and give time for students to explore these databases and tools on their own.

### REVIEWS/ASSESSMENT STRATEGY

The handout/worksheet associated with this lesson is intentionally not designed for library assessment purposes. It is instead intended to guide students to apply the concepts to their own research topic during class time (to “workshop” their topic) and then serve as reference notes as they pursue further inquiry on their own.

Ideally, students should have time and equipment during the workshop to search for data sets using recommended strategies and tools, with the librarian available for individual follow-up questions. Librarians can evaluate the effectiveness of their teaching based on these follow-up questions. As students learn to identify and access data in order to analyze them for a project such as an honors thesis or capstone paper, the true measure of their learning is their own successful completion of their work. No matter how successful this instruction session is, it is likely that students will individually consult with librarians and faculty throughout the life of their research project.

## Appendix A. Library Help for Data Research

Research Guide: [link to research guide](#)

Librarian: name and email

### DATA—DEFINED

**da·ta** noun plural but singular or plural in construction, often attributive \ˈdā-tə, ˈda- also ˈdä-\

1. factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation
2. information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful
3. information in numerical form that can be digitally transmitted or processed  
(From Merriam-Webster, <http://www.merriam-webster.com/dictionary/data>)

### TYPES OF DATA

- **Observational:** Captured in real-time, typically outside the lab  
Examples: Sensor readings, survey results, images, audio, video
- **Experimental:** Typically generated in the lab or under controlled conditions  
Examples: test results
- **Simulation:** Machine generated from test models  
Examples: climate models, economic models
- **Derived / Compiled:** Generated from existing datasets  
Examples: text and data mining, compiled database, 3D models

### COMMON FORMATS

- **Text:** field or laboratory notes, survey responses
- **Numeric:** tables, counts, measurements
- **Audiovisual:** images, sound recordings, video
- **Models, computer code, geospatial data**
- **Discipline-specific:** FITS in astronomy, CIF in chemistry
- **Instrument-specific:** equipment outputs

### KEY TERMS

- **Microdata:** Data directly observed or collected from a specific unit of observation
  - Examples
    - Census: the unit of observation is probably an individual, a household, or a family
    - Survey or poll: the responses of a single respondent
- **Aggregate Data:** Is higher-level data that have been compiled from smaller units of data
  - Examples: inflation rate, consumer price index, demographic data for city or state

- **Statistics:** Numerical data that has been organized and interpreted, usually displayed in tables
- **Datasets:** A dataset or study is made up of the raw data file and any related files, usually the *codebook* and *setup files*.
  - Most datasets require at least basic *statistical analysis* (Stata, SPSS, R, etc.) or *spreadsheet* programs (Excel) to use.
- **Repositories:** A *data repository* is a collection of datasets that have been deposited for storage and findability.
  - They are often discipline specific and/or affiliated with a research institution
  - Examples
    - ICPSR
    - Harvard Dataverse Network
    - UC San Diego Digital Collections

**Step by Step Plan for Finding Datasets and Statistics**

**1. Think about who might collect the data.**

- Could it have been collected by a government agency?
- A nonprofit/nongovernmental organization?
- A private business or industry group?
- Academic researchers?

Who? \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

**2. Look for publications that use the kind of data you’re looking for and cite the dataset**

- e.g. scholarly articles or government reports.
- **Quick tip:** use your article search strategies for Google Scholar and add the word “dataset”

Sample search: \_\_\_\_\_

**3. Once you know that what you want exists, it’s time to hunt it down.**

- Is it freely available on the web?
  - Google Dataset Search: <https://toolbox.google.com/dataset-search>
  - IQSS (Harvard) Dataverse
  - Quality of Government Institute (free)
  - MacroData guide
  - data.census.gov
  - UNdata
  - World Bank World Development Indicators
  - **Tip:** Check regular Google—you never know!
- Or part of a package to which the library already subscribes?

<ul style="list-style-type: none"> <li>• ICPSR (requires reg w/ UCSD email)</li> <li>• UC San Diego Dataverse</li> <li>• Cross-National Time Series</li> </ul>	<ul style="list-style-type: none"> <li>• CQ Voting &amp; Elections Collection</li> <li>• Roper iPoll</li> <li>• Proquest Statistical Insight</li> <li>• Data-Planet</li> <li>• OECD iLibrary</li> </ul>
--	---

- Can it be requested directly from the researcher? There’s a reason articles usually include author contact information...

