CC BY NC SA

About Codebooks

What is a codebook?

A codebook describes and documents the questions asked or items collected in a survey. Codebooks and study documentation will provide you with crucial details to help you decide whether or not a particular data collection will be useful in your research. The codebook will describe the subject of the survey or data collection, the sample and how it was constructed, and how the data were coded, entered, and processed. The questionnaire or survey instrument will be included along with a description or layout of how the data file is organized. Most codebooks are available electronically, and you can read them on your computer screen, download them to your machine, or print them out.

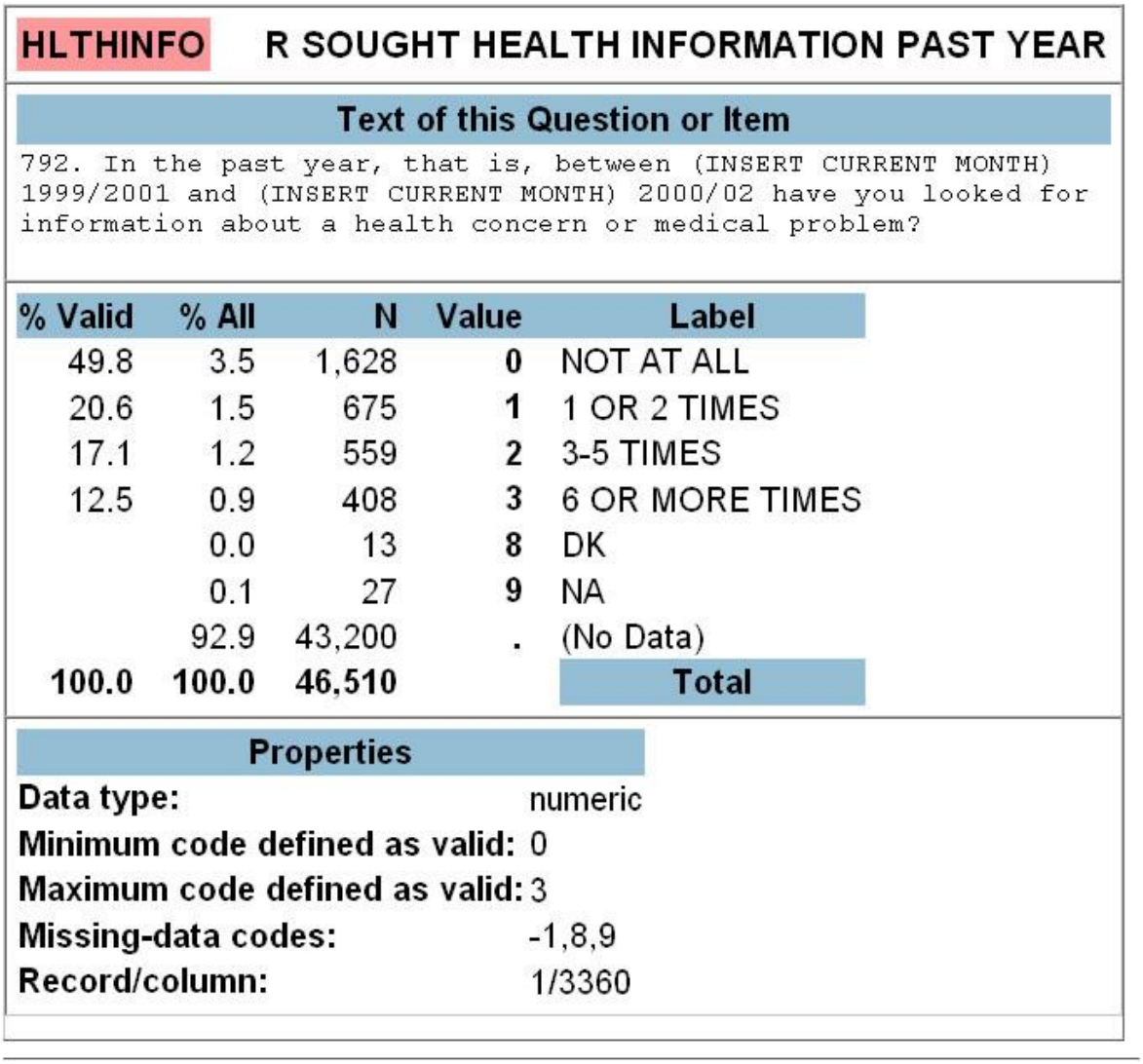
SSDA Archive Policy on Codebooks

Most codebooks in the Archive are machine-readable, and can be downloaded from our web site, or from our Dataverse site. If the data were acquired from ICPSR, and you are a UCLA student, faculty or staff member, you can go to the ICPSR web site, and download the codebooks from there. Codebooks are also available for polls available from the Roper Center for UCLA faculty, staff and students.

Codebook Example

On the next page is a diagram of the details included in most codebooks. This is a simple example of a codebook. In a complex survey, there will be more details about the flow of questions asked and the electronic organization of the data. In addition to question text, the most important items are variable names, values, value labels, and column locations. Try to find them in the example below and familiarize yourself with the layout of a codebook.

CC BY NC SA



HLTHINFO R SOUGHT HEALTH INFORMATION PAST YEAR

Text of this Question or Item

792. In the past year, that is, between (INSERT CURRENT MONTH) 1999/2001 and (INSERT CURRENT MONTH) 2000/02 have you looked for information about a health concern or medical problem?

% Valid	% All	N	Value	Label
49.8	3.5	1,628	0	NOT AT ALL
20.6	1.5	675	1	1 OR 2 TIMES
17.1	1.2	559	2	3-5 TIMES
12.5	0.9	408	3	6 OR MORE TIMES
	0.0	13	8	DK
	0.1	27	9	NA
	92.9	43,200	.	(No Data)
100.0	**100.0**	**46,510**		**Total**

Properties

Data type:	numeric
Minimum code defined as valid:	0
Maximum code defined as valid:	3
Missing-data codes:	-1,8,9
Record/column:	1/3360

Here is an explanation of the codebook image above. The circled red numbers relate to each of the subject headings below.

1) Item or Variable Name: This is usually a mnemonic, or nickname assigned to an individual question.
2) Variable Label: A short summary or description of question content.
3) Card and Column Locations: Indicates the electronic location of numerically coded responses to questions.
4) Question Text: Exact text of a question as delivered to a respondent.
5) Values and Value Labels: Describes the numeric and textual response options or categories to questions.
6) Valid Responses: Indicates the allowed numeric codes to question responses.
7) Branching: Indicates the flow of the questionnaire.