Organizing Data in Tables and Charts: Different Criteria for Different Tasks

KEYWORDS:
Teaching;
Charts;
Data presentation;
Empirical;
Tables;
Theoretical.

Jane E. Miller
Rutgers University, New Brunswick, New Jersey, USA.
e-mail: jem@rci.rutgers.edu

Summary
Tables and charts are efficient tools for organizing numbers, but many people give little consideration to the order in which they present the data. This article illustrates the strengths and weaknesses of four criteria for organizing data – empirical, theoretical, alphabetical and a standardized reporting scheme.

◆ INTRODUCTION ◆

Tables and charts are efficient tools for organizing numbers. Too often, however, students and quantitative analysts do not give much consideration to the order in which they present data in tables or charts. This lack of thought means that the sequence of items may not be compatible with the author's objectives, whether testing a hypothesis, describing a pattern or reporting data for others' use. The appropriate criteria for arranging data in tables or charts often differ depending on whether they are to be used primarily with or without a prose description. Tables or charts intended to present numbers as evidence to address a specific question or to accompany a description of a pattern are usually best organized so that they coordinate with the associated narrative. On the other hand, tables intended to present data for reference use such as periodic series from the Bureau of Labour Statistics or a national census might work better if structured so that readers can find the numbers of interest to them with little written guidance. These two broad objectives suggest very different considerations for organizing variables or response categories.

No one of these purposes is inherently more important than the other, but in many cases a particular objective can be identified for reporting numbers in a given type of document. For example, numbers presented in a science laboratory report or a history essay are usually being applied as evidence for a particular hypothesis or to illustrate a trend or other pattern. In such cases, empirical or theoretical criteria are frequently a sensible basis for arranging the data because that is how they will be discussed in the accompanying prose. In contrast, detailed reference data on population or income for each of a dozen or more dates or places might not come with a written description, so using a self-guiding convention is well-suited to such tasks. This article uses data from the U.S. Consumer Expenditure Survey to illustrate four approaches to organizing data within tables and charts, discussing the situations for which each approach might be preferred.

Table 1 presents data on major categories of expenditures from the 2002 Consumer Expenditure Survey (CEX). The CEX is conducted annually by the Bureau of Labor Statistics (BLS) using a diary survey form to collect detailed information on expenditures (U.S. Department of Labor 2004a). The information is then coded into the standard categories shown in table 1, which retains the original order of major expenditure categories from a standard BLS report (U.S. Department of Labor 2004b).

◆ ORGANIZING DATA TO ACCOMPANY A PROSE DESCRIPTION ◆

When testing hypotheses or portraying trends or other patterns, it is helpful to organize your data in tables or charts in the order you will describe them. For such purposes, alphabetical order and

Item	*Expenditures ($)*
Average annual expenditures	42,557
Food	5612
Alcoholic beverages	415
Housing	13,481
Apparel and services	1872
Transportation	7984
Health care	2410
Entertainment	2167
Personal care products and services	562
Reading	145
Education	771
Tobacco products and supplies	334
Miscellaneous	846
Cash contributions	1366
Personal insurance and pensions	4593

Table 1. Average annual expenditures by major expenditure category,[a] U.S. Consumer Expenditure Survey, 2002

Source: U.S. Department of Labor, Bureau of Labor Statistics, 2004a. Table 1.

[a]For all households with complete income reporting.

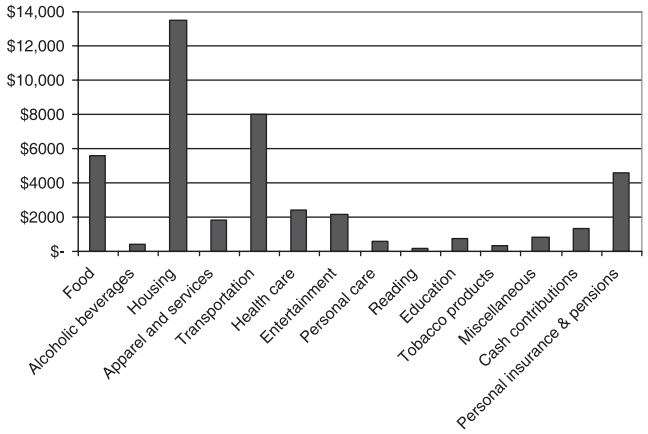

Fig. 1. Major categories of expenditures, BLS ordering, 2002 U.S. Consumer Expenditure Survey

the sequence of items from the original data source are poor organizing principles because they rarely correspond to substantively interesting or empirically relevant patterns.

Consider figure 1, which presents the information from table 1 in chart form, again preserving the order of expenditure categories from the BLS report. The heights of the bars and the conceptual content of adjacent categories vary erratically, requiring readers to zigzag back and forth across the axes to identify the rank order of expenditure categories by dollar amount or to compare categories of necessities to one another or to non-necessities.

For similar reasons, figure 2 – which sequences the expenditure categories alphabetically – would also be a poor choice to accompany a description of empirical rankings or a discussion of necessities versus non-necessities. Instead, to complement a

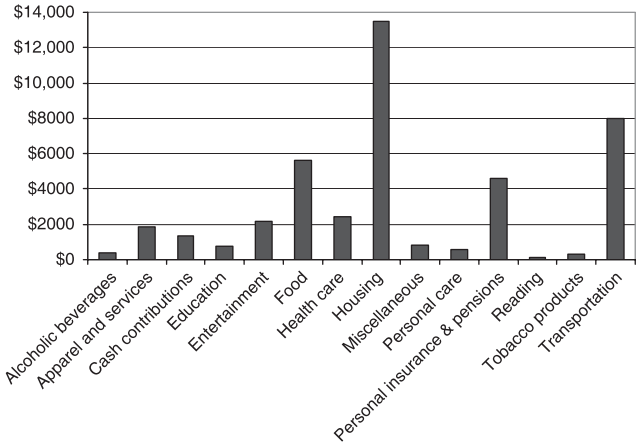

Fig. 2. Major categories of expenditures, alphabetical order, 2002 U.S. Consumer Expenditure Survey

prose description of a pattern, it is often sensible to arrange your data so that the audience can easily follow the associated narrative using the well-established conventions of tracking left-to-right and top-to-bottom within the table or chart. Before creating the table or chart or writing the associated prose, consider which of the organizing principles described below best matches the main point you wish to make. Arrange the rows and columns (or axes and legend) accordingly, and then describe the numbers in the same order as they appear in the table or chart.

Which organizing criterion to use depends largely on the type of variables in question. When reporting results for ordinal variables such as age group or income quintile, the sequence of items in rows, columns or axes will be obvious. Likewise, it makes sense to follow the natural order of values for interval or ratio variables such as date, age in years or income in dollars or pounds.

For tables or charts that present nominal variables such as favourite flavour of ice cream, or items such as categories of consumer expenditures, the categories or variables lack an inherent order. In those instances, either empirical criteria or theoretical principles usually provide a good basis for deciding how to organize them.

Empirical ordering

For many tables or charts presenting distributions or associations, an important aim is to show which items have the highest and the lowest values and where other categories fall relative to those extremes. If this is your main point, it is often suitable to organize the categories in ascending or

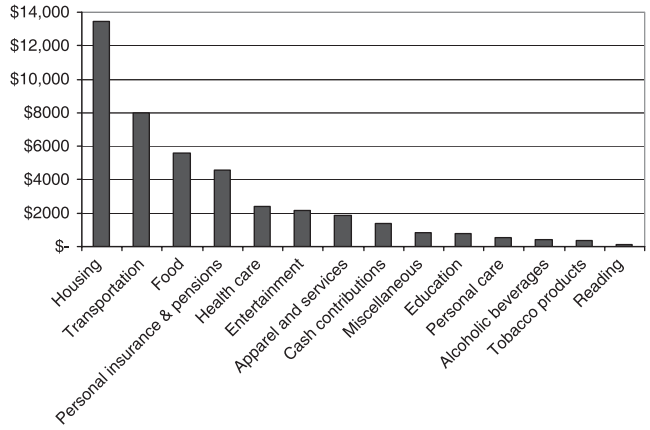

Fig. 3. Major categories of expenditures, descending dollar value, 2002 U.S. Consumer Expenditure Survey

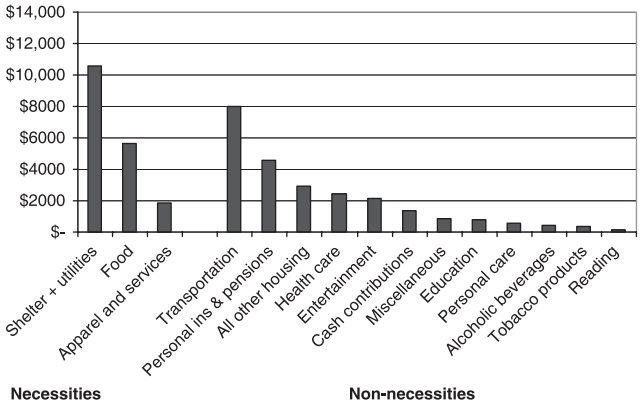

Fig. 4. Descending dollar value of expenditures, necessities/non-necessities, 2002 U.S. Consumer Expenditure Survey

descending order of frequency or value. For example, figure 3 shows major categories of consumer expenditures in descending order of dollar value.

Theoretical grouping

Arranging items into conceptually related sets can be very effective. For example, Duly (2003) reports statistics on consumer expenditures for necessities, which she defines as housing (including shelter and utilities but excluding other categories of housing-related expenses), food and apparel. To present the associated numbers, figure 4 groups the expenditure categories into necessities on the left-hand side of the *x*-axis and non-necessities on the right-hand side, with axis titles to identify those classifications. A table version would comprise separate panels for necessities and non-necessities, each with rows reporting the respective component categories. The accompanying description could then contrast the relative shares of necessities and non-necessities without requiring the audience to meander all over the table or chart to find the pertinent numbers.

Using multiple organizing criteria

For tables or charts that present more than a few variables, a combination of approaches is often useful. For instance, consider grouping items theoretically and then arranging them within those groups in order of descending frequency or other empirical consideration. Figure 4 divides categories of consumer expenditures into necessities and non-necessities, and then organizes them in descending order of dollar value within each of those classifications, providing a useful structure for pointing out key patterns in the data.

Sometimes it makes sense to apply the same criterion sequentially, such as identifying major theoretical groupings and then minor topic groupings within them. Among the necessity categories of consumer expenditures are items related to housing, food and apparel, each with a major heading. Within each of those major categories would be minor categories and subcategories, such as shelter and utilities as subcategories under housing.

For charts or tables organized into several theoretically or empirically similar groups of items, alphabetical order can be a logical way to sequence items within those groups. For example, data on all the nations of the world might be grouped by continents, and then listed alphabetically within each continent. Alphabetizing within conceptual or empirical groupings also works well if several items have the same value of the statistics reported in the table (e.g. mean value or frequency).

Writing a narrative to accompany the table or chart

Having created a table or chart that presents data in empirical or theoretical order, it is usually helpful to write the narrative to coordinate with that pattern, mentioning the organizing principle as you refer to the associated table or chart. For example, to describe the empirical pattern across categories of consumer expenditures, you might write:

> 'Figure 3 presents average consumer expenditures for the United States in 2002 in descending order of dollar value. Housing was the highest single highest expenditure category, followed by transportation, food and personal expenditures and pensions . . .'

An analysis that compares necessities and non-necessities could read:

'Figure 4 shows average consumer expenditures for necessities and non-necessities in the U.S. in 2002. Among necessities, shelter was the highest. . . . Among non-necessities, transportation . . .'

◆ ORGANIZING DATA FOR REFERENCE USE ◆

Reference documents typically include little if any prose description, so using a familiar convention or standard sequence is a sensible way to help readers find specific information quickly.

Alphabetical order

Alphabetical order is a widely understood organizing principle, commonly used in a variety of settings. For example, the daily stock market report of opening, closing, high and low prices effectively organizes thousands of numbers in a predictable format that readers can use without guidance.

Order of items from a standard document

Reference data from periodic surveys, censuses or surveillance systems are frequently best organized using the order of items from the original data collection instrument or following a standard coding or reporting scheme for that data source. People wishing to use those data often locate the variables of interest using original documentation such as the code book, survey instrument or census form, or by consulting copies of earlier volumes of the same reference publication. Using that standardized approach to organize tables or charts of reference data facilitates users' collection efforts by maintaining consistency across sources. Table 1 above employs such an approach, presenting the expenditure categories in the standard order used in many BLS reports (U.S. Department of Labor 2004b). Having collected the data of interest to them, users can then organize those numbers to suit to their objectives, whether summarizing trends for a report or testing a hypothesis about relationships among variables.

◆ SUMMARY ◆

Using appropriate organizing principles can significantly enhance the efficacy of a quantitative description or increase the accessibility of reference data. Readers of this article who would like to go further into the issues explored here may find it useful to refer to Miller (2004).

References

Duly, A. (2003). Consumer spending for necessities. *Monthly Labor Review*, **126**(5), 3. Available online at http://stats.bls.gov/opub/mlr/2003/05/art1full.pdf. Accessed June 2007.

Miller, J.E. (2004). *The Chicago Guide to Writing about Numbers.* The Chicago Guides to Writing, Editing, and Publishing. Chicago: University of Chicago Press.

U.S. Bureau of Labor Statistics (2004a). *Consumer Expenditure Survey, Diary Survey Form.* Available online at http://stats.bls.gov/cex/csx801p.pdf. Accessed June 2007.

U.S. Department of Labor, Bureau of Labor Statistics (2004b). *Consumer Expenditures in 2002.* Report 974. Available online at http://stats.bls.gov/cex/csxann02.pdf. Accessed June 2007.