CHAPTER 1

INTRODUCTION TO TEXT

Qualitative Data Analysis

Introduction
What Is Qualitative Data Analysis?
What Are Data and What Makes Them Qualitative?
About Numbers and Words
Research Goals
1. Exploration
2. Description
3. Comparison
4. Testing Models

Five Kinds of Qualitative Data
1. Physical Objects
2. Still Images
3. Sounds
4. Moving Images: Video
5. Texts

Further Reading

INTRODUCTION

This book is about ways to produce and analyze qualitative data in the behavioral and social sciences.

All sciences rely heavily on—and have well-developed methods for the analysis of—qualitative data. When ecologists pore over satellite images of the Earth’s surface, when astronomers listen to recordings of sounds from other galaxies, and when medical researchers listen to heartbeats, they are all looking for regularities in qualitative data. “Looking for regularities” is analysis. It’s the quintessential qualitative act, and it’s common to all traditions of scholarship across the humanities and the sciences.
WHAT IS QUALITATIVE DATA ANALYSIS?

Because of a quirk in the English language, the phrase “qualitative data analysis” is mischievously ambiguous. It can mean “the analysis of qualitative data” or it can mean “the qualitative analysis of data.” The confusion can be eliminated by distinguishing clearly between data and analysis. Figure 1.1 lays out the possibilities.

The top left cell, A, shows the qualitative analysis of qualitative data. Interpretive studies of texts, like transcriptions of interviews, are of this kind. Investigators focus on and name themes in texts. They tell the story, as they see it, of how the themes are related to one another and how characteristics of the speaker or speakers account for the existence of certain themes and the absence of others. Researchers may deconstruct a text, look for hidden subtexts, and try to let their audience know—using the power of good rhetoric—the deeper meaning or the multiple meanings in it.

The bottom right cell, D, refers to numerical or statistical analysis of numerical data. Lots and lots of data about human behavior come to us as numbers. Closed-ended questions in surveys produce numerical data. So do national censuses. Organizations, from businesses to charities to zoos, produce numerical data, too—data about the socioeconomic characteristics of people who use their products or services, data about how often they have to replace managers, data about how much time secretaries spend on the phone and on e-mail, and on and on.

**Figure 1.1** Key Qualitative and Quantitative Distinctions

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>A Interpretive text studies.</td>
</tr>
<tr>
<td></td>
<td>Hermeneutics, Grounded Theory, etc.</td>
</tr>
<tr>
<td>Quantitative</td>
<td>C Turning words into numbers.</td>
</tr>
<tr>
<td></td>
<td>Classic Content Analysis, Word Counts, Free Lists, Pile Sorts, etc.</td>
</tr>
<tr>
<td></td>
<td>D Statistical and mathematical analysis of numeric data</td>
</tr>
</tbody>
</table>

The top right cell, B, is the qualitative analysis of quantitative data. It’s what quantitative analysts do after they get through doing the work in the quantitative/quantitative cell, D, and it involves the search for, and the presentation of, meaning in the results of quantitative data processing. The qualitative/quantitative cell, B, includes everything from the finding of regularities in a scatter plot to the interpretation of meaning and substantive significance of statistical tests. Without the work in the qualitative/quantitative cell, the kinds of studies shown in the quantitative/quantitative cell are sterile and vacuous.

Which leaves the bottom left cell, C, the quantitative analysis of qualitative data. This involves turning words, images, sounds, or objects into numbers. Scholars in communications, for example, tag a set of television ads from Mexico and the United States to test differences in how older people are portrayed in the two countries. Political scientists code the rhetoric of a presidential debate to look for patterns and predictors of policies. Archeologists code a set of artifacts to produce emergent categories or styles or to test whether some intrusive artifacts can be traced to a source.

In this book, we’ll be mostly concerned with cells A (qualitative/quantitative), B (qualitative/quantitative), and C (quantitative/quantitative).

**WHAT ARE DATA AND WHAT MAKES THEM QUALITATIVE?**

Data—qualitative and quantitative alike—are reductions of our experience (Bernard et al. 1986). Electrons and DNA are things. With a little help from some instruments, we can look at electrons and DNA and we can record what we see. Whatever we choose to record about things like these— their shape, their size, their weight, their speed—are data. If we record numbers, we get quantitative data; if we record sounds, words, or pictures, we get qualitative data.

In the social sciences, we are interested in people’s behavior, thoughts, emotions, and artifacts (the physical residue of people’s thoughts, emotions, and behavior) and the environmental conditions in which people behave, think, feel, and make things.

When we reduce our experience of those things to numbers, the result is quantitative data. And when we reduce people’s thoughts, behaviors, emotions, artifacts, and environments to sounds, words, or pictures, the result is qualitative data.

We create data by chunking experience into recordable units. Consider three researchers observing children at play in a schoolyard or playground. One of them watches the children and writes up field notes on what she or he saw. Another records the frequency of particular behaviors using a checklist. The third uses a video camera to record the children playing.
phenomena of interest—the behavior, the words, the laughter, and the
crying of children on a playground—are ephemera, disappearing as they
happen. The records of the phenomena—the notes, the checklist, the video
recording—remain for us to analyze and understand. Data are the archeological
record of experience.

Some qualitative data are produced on purpose—we interview people
and transcribe their words; we put children together in a room full of toys
and videotape or take notes about what they do—but most of the record
about human thought and behavior comes to us as naturally occurring
qualitative data. The paintings produced during the first hundred years of the
Italian Renaissance, the television ads that aired last week in Mexico that
contained images of old people, the articles in the Wall Street Journal over
the last 20 years that contain the phrase “corporate culture,” the diaries of
U.S. Civil War soldiers, and the blogs of today’s soldiers in Iraq—all are
naturally occurring, qualitative data.

Across the sciences, from anthropology to zoology, from sociology to
physics, data—all data, qualitative and quantitative—are selections of what’s
available. Satellites don’t record everything going on below them any more
than observers of human behavior record everything they see. People—real
human beings—decide to measure some things and not others. These
decisions are not random. They are sometimes based on unadulterated
scientific curiosity, and they are sometimes based on what’s fundable. They
are sometimes motivated by humanitarian instincts, and sometimes
motivated by greed. This does not invalidate the effort to produce data. It
does, however, remind us that there is a human component to science, just
as there is in art, government, or commerce.

About Numbers and Words

Every reader of this book is aware of the longstanding debate in the social
and behavioral sciences about the relative merits of quantitative versus qual-
itative data. The debate reflects principled stands by colleagues who identify
with the positivist tradition and those who identify with the humanist tradi-
tion of research. These discussions about epistemology—how we know
things at all—have a noble lineage dating to Protagoras’ (485–410 BCE)
famous maxim that “man is the measure of all things”—meaning that truth
is not absolute but is decided by individual human judgment—and to
Lucretius’ (94–49 BCE) insistence on the material nature of all things, includ-
ing the mind. (Further Reading: the qualitative-quantitative issue.)

The qual-quant conflict plays out in all fields of social and behavioral
science. In psychology, most research is in the positivist tradition, but much
**Chapter 1  Introduction to Text**  ◆  7

*clinical work* is in the humanist tradition because, as its practitioners sensibly point out, it works. In cultural anthropology, data are collected by fieldworkers—which makes cultural anthropology thoroughly empirical—but much of the data *analysis* is done in the humanist or interpretivist tradition.

Most research in sociology today is positivist, tracing its lineage to August Comte, Adolphe Quetelet, and Emile Durkheim. The increasing number of sociologists today who count themselves as interpretivists, however, can trace their epistemological roots to the great tradition of Immanuel Kant, Wilhelm Dilthey, and others in the school of German Idealism. (Idealism here refers not to the pursuit of high purpose, but to the precedence of reason, or ideas, over empiricism in the practice of science.) The same can be said about research in education, nursing, and other fields: Positivists and interpretivists alike have long traditions on their side.

Notice that we don’t say anything like “Research in X is mostly quantitative” or that “Research in Y is mostly qualitative.” In fact, we never use the distinction between quantitative and qualitative as cover for talking about the difference between science and humanism or between interpretivism and positivism. Lots of scientists do their work without numbers, and many scientists whose work is highly quantitative consider themselves to be humanists as well. Moreover, numbers do not make an inquiry scientific—searching the Bible for statistical evidence to support the subjugation of women doesn’t turn the enterprise into science—and the use of qualitative data does not diminish the scientific credibility of any piece of research (see Box 1.1).

---

**Box 1.1**

**Numbers and Science: What’s the Fuss About?**

Scholars in the physical and biological sciences wonder what all the fuss is about. They already know how powerful qualitative data are. Satellite images inform geology, meteorology, astronomy, ecology, archeology, and oceanography. Images from electron microscopes inform chemistry, molecular biology, and physiology. Lengthy narratives dictated into a tape recorder by observers inform students of volcanoes, hurricanes, gorillas, and crime scenes.

Researchers in what we usually think of as highly quantitative sciences use a whole family of qualitative methods, called visualization methods, for understanding patterns in numerical data. Multidimensional scaling, for example, is a visualization method that’s widely used in the social sciences. In fact, it was developed in the social sciences but, like all useful methods, is used across all sciences now. More on multidimensional scaling in Chapters 5 and 8.
There are four main objectives in qualitative research, irrespective of whether it's based on qualitative or quantitative data. The questions associated with each are shown in Table 1.1.

1. Exploration

At this early stage, the goal is to discover themes and patterns and to build initial models of how complex systems work. Whether we’re talking about astronomers scanning the night sky in search of new comets and asteroids or grounded theorists studying how people experience illness, exploring means following leads and hunches; taking a step forward and then backtracking; uncovering what’s there; experiencing the phenomenon we’re studying, if possible; and identifying both its unique features and the features it shares with other phenomena.

<table>
<thead>
<tr>
<th>General Aim</th>
<th>Type</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exploration</td>
<td>Case</td>
<td>What kinds of things are present here?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How are these things related to one another?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Are there natural groups of things here?</td>
</tr>
<tr>
<td>2. Description</td>
<td>Case</td>
<td>What does a case look like?</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>What does a set of cases look like?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Is a particular kind of thing (A) present or not?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How much of that kind of thing (A) is there?</td>
</tr>
<tr>
<td></td>
<td>Cultural</td>
<td>What does the culture look like?</td>
</tr>
<tr>
<td>3. Comparison</td>
<td>Case</td>
<td>How is case X different from case Y?</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>How is a group of Xs different from a group of Ys?</td>
</tr>
<tr>
<td>4. Testing models</td>
<td>Case</td>
<td>To what degree does a particular case conform to the proposed model?</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>To what degree does a group of cases conform to the proposed model?</td>
</tr>
</tbody>
</table>
2. Description

Every field of science depends vitally on good description. Long before the physics of avian flight were worked out, people watched and recorded as faithfully as possible just how birds managed not to fall out of the sky. Every new comet and asteroid that’s discovered is described in the scientific literature, as is every new bug and plant and disease.

Descriptions can be qualitative or quantitative, or both, and detailed case studies—with their listings of typical features, idiosyncrasies, and exceptions—are used widely in the teaching of law, medicine, and management. Ethnographic field notes are typically filled with individual case studies.

How much precision should you shoot for in a good description? When you’re collecting data, you should get as much as possible. You can always back off on the level of precision later, when you write up your findings. It may be enough in your write-up to say something like “Cambodian refugees comprise the largest ethnic group in this neighborhood.” But if you need to know the percentage of each ethnic group in the neighborhood, then you’d better collect that data from the start by asking every refugee what his or her ethnicity (or language or country of origin) is. You can always generalize from specifics, but you can never go the other way.

In describing cultural beliefs and practices, we focus on what people share and what they don’t share. Here again, it pays to get as much as possible from the start—and for the same reason: You can only generalize if you have the specifics.

3. Comparison

Qualitative comparison involves identifying features that individuals or groups share and don’t share. In the 1930s, Wayne Dennis, a psychologist, collected observational data on 41 Navajo and Hopi babies and on a similar group of White American babies in a study of child-rearing practices. Here’s a thoroughly qualitative, comparative statement from his study: “Whereas some American infants are bottle-fed almost from the beginning and many are breast fed but a short time, all Hopi infants are breast fed, none are weaned under one year of age and many are not weaned before two years” (Dennis 1940:307).

Quantitative comparison involves testing whether (and how much) measurements of variables track each other. Does the weight of children between the ages of 10 and 16 vary with their height? If so, how closely do
the two variables (height and weight) track each other? Does the tracking vary by ethnic group? By family income?

Just as with description, it pays to collect specific data and not rush to generalize.

4. Testing Models

This is where we test hypotheses against observations. We can do this with only qualitative data, with only quantitative data, or with both. Here is Wayne Dennis again:

American infants are usually placed on a rigid time schedule of feedings with an interval of several hours between feedings [and are] . . . often expected to cry for a period before being fed. The Hopi infant, on the other hand, is nursed as soon as he cries, and consequently nurses frequently and cries very little. (Dennis 1940:307)

Dennis has drawn a strong conclusion about the relationship between crying and feeding—and has done so without reporting a single number. In fact, he was testing a much, much larger model, or set of hypotheses, about the care of infants. After reporting on Hopi, Navajo, and White American practices for carrying, feeding, and toilet training of infants, Dennis concludes that “beginning roughly at one year of age the patterns of the infant begin to vary in accordance with the culture of the group” as children begin to learn a language and to imitate their parents’ distinctly cultural behavior. Dennis concludes that “this corroborates the view that the characteristics of infancy are universal and that culture overlays or modifies a more basic substratum of behavior” (1940:316).

Many projects involve all four of these activities—exploration, description, comparison, and model testing. Some scholars rely on qualitative data for exploration and discovery and rely on quantitative data for testing models. Increasingly, though, research across the social sciences relies on a balanced, commonsensical mix of both kinds of data.

♦ FIVE KINDS OF QUALITATIVE DATA

Qualitative data come to us in five forms: physical objects, still images, sounds, moving images and, of course, written words (see Box 1.2).
Table 1.2 shows the five kinds of qualitative data, broken down by size and accessibility. Material objects range from personal trinkets to vast remains of ancient cities. Videos can be 30-second commercials or 3-hour epic pictures. Graphic data include stick figures drawn by children and murals by Diego Rivera that cover entire museum walls. Texts can be single-word answers to questions or the complete works of Shakespeare or transcribed narratives from ethnographic interviews. Data from public sources are more accessible than are data from private ones.

1. Physical Objects

For archeologists who study preliterate societies—societies that flourished before written communication—physical remains may be the only data available. The study of material culture, however, is not limited to societies of the distant past. Beginning in the late 15th century, the Age of Discovery in Europe produced an enormous market for material objects from societies around the world and by the late 19th century, anthropologists in Germany, Britain, and the United States were avid collectors of artifacts for public museums.

*Notes and Queries on Anthropology*, the methodological bible for anthropologists up to about 50 years ago, was compiled by the Royal Anthropological Institute of Great Britain. 

---

**Box 1.2**

**Measuring Smells and Tastes**

Interestingly, there are very limited data forms for taste or smell. We know how important these *phenomena* are to people—try to imagine ethnicity in the United States without referring to burritos and lasagna and bagels and moussaka and pirogis. Psychologists and chemists are working on ways to measure taste and smell but, so far, the only easily accessible olfactory and gustatory *data* we have about these things come from our memories or from self-reports. Wine aficionados have developed an elaborate vocabulary about tastes and smells—a vocabulary that turns memory into exchangeable information.
Table 1.2 Kinds of Qualitative Data Based on Form, Size, and Accessibility

<table>
<thead>
<tr>
<th>Form</th>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accessibility</td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>Public</td>
<td>Public</td>
</tr>
<tr>
<td>Private</td>
<td>Personal jewelry, pill bottles, blood samples</td>
<td>Archaeological ruins, buildings, houses, universities, skyscrapers</td>
</tr>
<tr>
<td>Public</td>
<td>Park sculptures, street signs, pottery shards, store merchandise</td>
<td>Household garbage, clothing</td>
</tr>
<tr>
<td>Private</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Objects</td>
<td>Doodles, line sketches, family portraits, patient X-rays</td>
<td>Large detailed murals, art exhibits</td>
</tr>
<tr>
<td>Still Images</td>
<td>Magazine ads, cave art, billboards, Web pages, paintings hung in museums</td>
<td>Family albums, art portfolios, CAT scans</td>
</tr>
<tr>
<td>Sounds</td>
<td>Jingles, radio ads, intercom announcements, messages you hear while on hold</td>
<td>Political speeches, sports play-by-plays, music albums, focus group tape recordings</td>
</tr>
<tr>
<td>Moving Images: Video</td>
<td>TV ads, news footage, sitcoms</td>
<td>Full-length movies, documentaries, television programs</td>
</tr>
<tr>
<td>Texts</td>
<td>Epitaphs, obituaries, personal ads, political buttons, parking tickets</td>
<td>Thank-you letters, shopping lists, short responses to interview questions, e-mails</td>
</tr>
<tr>
<td></td>
<td>Books, manuals, religious tomes, court transcripts, Congressional Record, newspapers</td>
<td>Diaries, detailed correspondence, private chat-room discussions</td>
</tr>
</tbody>
</table>

devoted more than a hundred pages to the study of material culture. A sample:

The study of all aspects of the material side of people’s life is of great interest and importance not only from the intrinsic interests of the
artefacts themselves, but for sources of invention, and questions of diffusion. Further, artefacts and techniques have great importance by virtue of their relation to the whole social organization and to religious and other ceremonial practices. (Royal Anthropological Institute 1951:221)

Today, museums around the world provide us with a living record of the diversity of human religious, political, and economic activity, and social scientists from many disciplines continue to study human interaction with material objects.

Researchers in marketing and consumer behavior are vitally interested in material culture (D. Miller 1987; Therkelsen and Gram 2008). Students of the world’s religions collect and analyze icons and talismans (Handloff 1982; McColl 1982). Anthropologists have long used material possessions as indicators of status, prestige, and wealth in a community (B. R. DeWalt 1979). Ryan (1995), for example, found that the presence or absence of certain material objects—things like toilets, televisions, cars, corn mills, lamps—in the homes of African villagers predicted the kind of medical treatment that people sought. (Further Reading: material culture and museums.)

2. Still Images

For art historians, media and communication specialists, and for those who study popular culture, images, both still and moving, are standard forms of data. As early as 1919, Alfred Kroeber analyzed pictures in American and French fashion magazines and found “an underlying pulsation in the width of civilized women’s skirts, which is symmetrical and extends in its up and down beat over a full century; and an analogous rhythm in skirt length, but with a period of only about a third the duration” (Kroeber 1919:257).

Since then, there have been hundreds of social science studies using still images—greeting cards (Bridges 1993), comic strips (LaRossa et al. 2000), pictures in ads (Goffman 1979), photographs (Drazin and Frolich 2007)—as basic data. Malkin et al. (1999) analyzed the covers of 12 popular women’s magazines (Ladies Home Journal, Cosmopolitan, etc.) and nine popular men’s magazines (Esquire, Sports Illustrated, etc.). The culturally patterned messages are clear: Men are enjoined to expand their knowledge, hobbies, and activities; women are enjoined to improve their life by losing weight and doing other things to change their appearance.
3. Sounds

Audio data include things like music, narratives, speeches, radio programs, and taped interviews. Alan Lomax (1977), an ethnomusicologist, analyzed speech samples from 114 societies around the world. He found that there are regularities in speech styles in complex, midlevel, and primitive economies. For example, the length of spontaneous utterances is longer, on average, in more complex societies than in less complex ones. Barbara Ayres (1973) discovered a strong preference for systematic, repetitive rhythms in societies where infants are carried in slings and shawls. In societies where infants are rocked in cradles, there is a preference for irregular rhythms.

Sociolinguists and scholars of discourse are dedicated users of audio data. Labov and Waletzky (1997), for example, used tape-recorded narratives to understand differences in class and ethnic markers of Black and White American speech.

Discourse analysis is widely used in the study of human interaction—doctors and patients during examinations, pupils and teachers in classrooms, husbands and wives during counseling sessions, and so on. Part of the content of these interactions—a large part—shows up in basic audio transcriptions. But part of it—the part having to do with tone of voice, pitch, cadence, rhythm—doesn’t, and that part may be crucial to understanding what’s going on.

For example, Joel Sherzer (1994) compared a tape-recorded, 2-hour traditional chant by Chief Olopinikwa of the San Blas Kuna Indians in Panama, with a phonetic transcription of the event. The transcription left out the chanted utterances of the responding chief (usually something like “so it is”), which was key to understanding the verse structure of the chant. This may seem like an exotic example, but it isn’t. It’s just an example from an exotic language. In fact, prosodic markers (things like tone of voice, etc.) tell a lot of the story when we want to know the full meaning of an utterance or a piece of dialog. (Further Reading: linguistic analysis of discourse.)

4. Moving Images: Video

Moving images, or what we call video in the rest of this book, combines the power of images and sounds through time. Scholars in film studies,
sociologists, political scientists, and researchers in gender and media studies have become expert in analyzing video documents such as films, television programs and commercials, political ads, and even home-made movies.

Cowan and O’Brien (1990), for example, studied 474 cases of victims in slasher movies. Most protagonists in those movies are killed (that’s the whole point of the genre), but a few survive. Women who survive, it turns out, are less physically attractive than nonsurviving women and are not associated with any sexual behavior. The male nonsurvivors were cynical, egotistical, and dictatorial. Cowan and O’Brien conclude that, in slasher films, sexually pure women survive and “unmitigated masculinity” ends in death (1990:195).

The Third International Mathematics and Science Study, or TIMSS, was a massive study of how math and science were taught in 41 countries around the world in the 1990s. One part of the TIMSS effort was the intensive study of instructional practices and lesson content in three countries: Japan, Germany, and the United States. Researchers studied videotapes of eighth-grade classrooms in the three countries and found very different teaching styles. In the United States and Germany, students spent nearly all their time practicing routine procedures to learn math. In Japan, students spend less than half their time on this kind of learning and a lot of time figuring out new solutions to standard problems—an effort that stimulates conceptual, rather than rote thinking about mathematics (Jacobs et al. 2007; Stigler et al. 1999:vii). (Further Reading: images as qualitative data.)

5. Texts

By far the largest trove of qualitative data is the mountain of written texts that have been produced over the centuries. Scholars from the humanist and positivist traditions alike rely on texts as their primary data. Folklorists, sociologists, psychologists, anthropologists, and political scientists have analyzed newspaper articles, novels, congressional reports, brochures published by hate groups, personal want ads, court records, diaries, and, more recently, e-mail messages and web pages.

Most of this book is about analyzing this kind of qualitative data, but almost everything we have to say about finding themes, coding themes, and analyzing text can be applied as easily to objects, images, and sounds as they can to words.
Further Reading


♦ For more on the study of modern material culture, see the Journal of Material Culture, the Journal of Social Archaeology, and the Journal of Consumer Culture. See Dant (2005, 2006) for an overview. On clothing and fashion, see Crane and Bovone (2007). See Haldrup and Larsen (2006) on the importance of material objects in the study of tourism. For example, Holly and Cordy (2007) analyze the detritus left by visitors to gravesites as a way to document behavior (like vandalism, magic, legend tripping, and partying) that would be difficult to observe directly without long-term participant observation research.

♦ On material culture as a reflection of gender roles, see Chaterjee (2007). See Cavanaugh (2007) on how the production of a particular kind of food became a symbol for a town in Italy. See Öztürkmen (2003) for how material artifacts are used in the creation of nostalgic narratives about the past.

♦ For more on the role of museums and their artifacts in shaping culture, see Coombe (1994), Hilden and Huhndorf (1999), and P. M. Taylor (1995).

♦ For more on linguistic analysis of discourse, see Drew and Heritage (2006), Schegloff (2007), and Wennerstrom (2001), and see Chapter 10 on conversation analysis. Scholarly journals that focus on this include the Journal of Pragmatics and Text and Talk.