CHAPTER 3

Data Collection in Geography

Overview

Learning Objectives:

- What is the distinction between primary and secondary data sources?
- What are the five major types of data collection in geography?
- What are some of the ways geographers and others have made a distinction between quantitative and qualitative methods, and how do they relate to scientific and humanistic approaches in geography?

In the previous chapter, we explained that the empirical part of scientific research involves systematically observing cases in order to record measurements of variables that reflect properties of those cases. Researchers analyze the resulting set of data (usually numbers) graphically, verbally, and mathematically in order to learn something about the properties of the cases. Data collection efforts do not generally go on continuously but are grouped into periods of activity focused on particular research issues or questions. Such a focused period of data collection and analysis is a **study** (in Chapter 7, we learn that there are two major categories of scientific studies, experimental and nonexperimental). In this chapter, we introduce some basic characteristics of data collection in geography, including the distinction between primary and secondary data sources, the five major types of data collection, and the distinction between quantitative and qualitative methods.

Primary and Secondary Data Sources

One way to characterize data in geography concerns whether they were collected specifically for the purpose of a researcher's particular study. If so, we call the data **primary**. An example would be a geographer who interviews people about their attitudes toward bioengineered agriculture. If, instead, the data have been collected for another purpose, usually by someone other than the researcher, we call it **secondary**. An example of that would be a geographer who uses *Landsat* imagery to study landslides on the California coast. The imagery was not collected by that researcher, and it was not collected primarily so he or she could study landslides.

The major asset of primary data is that they are collected in a way specifically tailored to a particular research question, which means they are probably the data best suited to answering that question. In our attitude example above, the geographer would design the survey specifically to address the issue of attitudes toward bioengineering and agriculture, including customizing it to fit the people answering the survey and the place where they live. But all of this takes considerable time and effort to do well. In contrast, the major asset of secondary data is that they are sometimes the only data available to address a particular research question that are even moderately suited to that question. Also, secondary data are almost always less expensive than primary data (in terms of money, time, and effort). In our landslide example, the geographer gets a very large amount of free data obtainable in something like an hour or less, depending on the geographer's units of analysis, but that geographer has to accept the way the Landsat satellite collects imagery. This includes the extent of earth surface coverage, the time the satellite passes over, the spatial resolution of the imagery, and the spectral bands recorded.

Some geographers use mostly primary data, whereas others use mostly secondary data. This depends mostly on the geographer's topical area of research. However, compared to many other scientific disciplines, both human and physical geographers use a great deal of secondary data. This is probably because they so often study phenomena at large spatial and temporal scales, where it is typically so difficult and costly to collect data that a single study does not warrant it. The fact that secondary data are not tailored to the geographer's specific research question influences the nature of many geographers' research. Problems addressed by census data, for example, are the subject of more geographic research than is necessarily warranted from an intellectual or applied perspective. Especially characteristic of much geographic research in this respect is that researchers study problems at the analysis scale of the available data set, which is often not exactly the scale at which the phenomena operate (see Chapter 2). We consider the characteristically geographic problem that results from this "data-driven" approach to science several times in the rest of the text but especially in Chapter 9.

Types of Data Collection in Geography

We can characterize data in geography more precisely than just distinguishing primary from secondary. Geographers collect and analyze many different kinds of data

Table 3.1 Types of Data Collection in Geographic Research

1. Physical measurement (Chapter 4)
2. Observation of behavior (Chapter 5)
3. Archives (Chapter 5)
4. Explicit reports (Chapter 6)
5. Computational modeling (Chapter 7)

in their studies, based on many different variables and collected with many different techniques. However, we can group all of these data collection methods into just five types (Table 3.1). The first is very popular in geography, especially physical geography. **Physical measurements** consist of data collected by recording physical properties of the earth or its inhabitants. Physical properties include size and number, temperature, chemical makeup, moisture content, texture and hardness, the reflectance and transmissivity of electromagnetic energy (including optical light), air speed and pressure, and more. We discuss these physical measurements at some length in Chapter 4. One of the key innovations of 20th-century geography is the use of aerial and satellite remote sensing as ways to efficiently record large amounts of physical measurement data. We discuss physical measurement via remote sensing in Chapter 12. Human geographers often observe the "physical traces" left behind by human behavior or activity (biogeographers might study the physical traces of nonhuman animal activity). They include the house designs in different neighborhoods or cultural regions, crops that have been planted in different fields, or patterns of clear-cut forests left by different harvesting techniques.

The second type of data collection is based on the fact that human geographers also observe and record human behavior directly (again, biogeographers can observe animal behavior). **Behavior** is the overt and potentially observable actions or activities of individuals or groups of people. It is not their thoughts, feelings, or motivations, although very often behavioral observations provide the data that allow geographers to study thoughts, feelings, and motivations scientifically. Geographers make **behavioral observations** in person or with the aid of a variety of recording media. Importantly, records of behavior do not in themselves constitute data; they must be "coded" into categories to become data. As we discuss in Chapter 5, behavioral observations vary greatly in the degree to which they involve people's explicit awareness that they are being studied. We also consider in that chapter the important fact that people's behaviors are not always based on their explicit choices and decision-making.

A third type of data collection practiced by geographers is the use of existing records that others have collected primarily for nonresearch purposes, at least not the geographer's research. These secondary records are known as **archives**. Examples of archives used by geographers include financial records, birth and death records, newspaper stories, industry and business records, museum records, historical documents, diaries, letters, and more. Often, archives must also be coded in

order to produce usable data; for this reason, we discuss them in Chapter 5 along with behavioral observations.

The next type of data collection is quite popular in human geography. **Explicit reports** are beliefs people express about things—about themselves or other people, about places or events, about activities or objects. Actually, explicit reports are also observations of behavior; answering a question on a survey is behaving, for instance. But we distinguish reports as distinct types of data collection because they always involve explicit recognition by people that researchers are studying them, and because research participants' explicit beliefs and choices determine the data collected with explicit reports. Explicit reports such as surveys and interviews often consist of questions that have no right or wrong answers, or at least the correctness of the answer is not of chief interest to the researcher. When the explicit report consists of questions that do have right or wrong answers, and the correctness of answers *is* of interest to the researcher, we call the explicit report a test. That is, whereas many types of explicit reports are used to study opinions, attitudes, and preferences, tests are used to study knowledge. These measures are called "explicit" reports because people responding to them *know* they are responding to a request for information by a researcher. This turns out to be both an important strength and an important limitation of explicit reports, as we discuss in detail in Chapter 6.

The fifth and final type of data collection is **computational modeling**, applied in both physical and human geography. In Chapter 2, we defined models as simplified representations of portions of reality. We noted that models can be realized in conceptual, physical, graphical, or computational form. Understood in this broad way, models and modeling are pervasive in geography and other sciences. We refer to them frequently in this book, in several different chapters. For instance, in Chapter 9 we discuss statistical models, and in Chapter 10 we discuss graphical models (maps are models). We consider conceptual models in several different chapters, at least implicitly. As a unique approach to data collection, computational modeling is modeling that evaluates theoretical structures and processes expressed mathematically, typically in a computer. We discuss computational modeling in detail in Chapter 7, which covers research designs, because we believe it makes sense to think about modeling as an alternative to standard experimental and nonexperimental approaches. We see in Chapter 7 that we evaluate how well models fit portions of reality by comparing outputs of the model to measurements made on the reality to which the model refers. Alternatively, models are sometimes created and thought about as if they were creations of new realities rather than simulations of existing realities. We consider how this creation of "artificial realities" may or may not be thought of as *scientific* research in Chapter 7.

An Introduction to Quantitative and Qualitative Methods

Geographers, and other natural and social scientists, have been collecting and analyzing all of the types of data we have just discussed for well over a century

(of course, many specific techniques and procedures are regularly introduced). Besides geographers, these scientists have included geologists, biologists, oceanographers, hydrologists, atmospheric scientists, anthropologists, psychologists, sociologists, economists, political scientists, and others. Many of these early scientists incorporated a variety of data collection techniques and a variety of data types in order to understand their phenomena of interest. In other words, early scientists of the earth and its people were unabashedly heterogeneous in their empirical methods, using whatever they thought provided insight into their problem domain. We enthusiastically believe that this heterogeneous approach is still the best approach.

During the middle and latter part of the 20th century, characteristics of the varied methodological approaches applied in the sciences, particularly the social and behavioral sciences, were summarized in terms of a distinction between **quantitative** and **qualitative methods**. Like our definition of science in Chapter 1, the quantitative/qualitative distinction is difficult to define in a precise way. The distinction reflects a continuum as much as two sharp categories. There are clear examples of each but also examples that are more-or-less quantitative or qualitative.

A few different factors have been identified that distinguish quantitative and qualitative methods. One concerns the nature of the data recorded and analyzed in a research study. Quantitative data consist of numerical values, measured on at least an ordinal level but more likely a metric level. Qualitative data are nonnumerical, or, as in nominal data, numerical values that have no quantitative meaning. They consist of words (in natural language), drawings, photographs, and so on.

However, the distinction between quantitative and qualitative methods is not just whether a researcher uses numbers or not. Another factor distinguishing the two emphasizes the data collection technique used to create the data, rather than the data itself. According to this, quantitative methods are those that impose a relatively great amount of prior structure on collected data. That is, such methods involve a prior choice of constructs to study, a prior choice of variables with which to measure those constructs, and prior numerical categories with which to express the measured values of those variables. Qualitative methods, in contrast, involve less prior structure on data collection. Data collection that is very clearly qualitative might start with little more than a topic area or a broad research question. The constructs, variables, and especially the measurement values for the variables are determined as observations are made or even afterward. For example, a survey that asks respondents to pick one of a finite number of predetermined categories as a way to measure their attitudes about highway construction would be relatively quantitative in this sense; an interview that asks respondents "how they feel" about highway construction, without any constraints on what they can give for an answer, would be relatively qualitative. Importantly, these examples also show that a single type of data collection, in this case explicit reports (Chapter 6), may be used in a relatively quantitative *or* qualitative way.

Still another factor in differentiating quantitative and qualitative methods focuses on the analysis of data. Either methodological approach may start with relatively unstructured and open-ended responses, such as oral responses in an interview. These can be treated quantitatively, however, by rigorously **coding** the

elements in the responses (such as words or phrases) into well-defined categories, which are then tabulated and analyzed statistically; we detail this process in Chapter 5. Alternatively, some researchers choose to treat such records more qualitatively by avoiding rigorous coding and statistical analysis, instead interpreting the records more informally and in a less repeatable manner. Such a qualitative approach usually involves less aggregation of data as well. We return to these issues in Chapter 7, where we contrast single-case and multi-case research designs and further consider the nomothetic/idiographic distinction we touched on in Chapter 1.

We agree with researchers who endorse a "full spectrum" approach that incorporates multiple methods, including both relatively qualitative and quantitative methods. In many research contexts, the two complement each other rather well. Qualitative methods are more applicable when a researcher does not know much to begin with about a particular research topic or domain (that is, no research has been done on it). Such methods are flexible for the researcher, requiring less prior understanding of a phenomenon. Qualitative methods are especially useful in some areas of human geography because they allow a research case that is a sentient human being to "speak in his or her own voice," focusing on what is meaningful or important to himself or herself rather than conforming to the researcher's conceptualization of a situation. However, if qualitative methods are to provide scientifically acceptable evidence about the status of hypotheses and theories, they require difficult and laborious coding. Among other things, this means that researchers working qualitatively typically examine fewer of the cases of interest. Both the relatively low reliability of coding and the typically smaller samples of cases mean that qualitative methods generally do not produce evidence that can be as convincingly generalized as that produced by quantitative methods. For these reasons, we generally favor a research strategy that applies qualitative approaches earlier in an exploratory way and quantitative approaches later, informed by the qualitative results, in a confirmatory way. In many areas of human geography, furthermore, qualitative results can help flesh out and exemplify quantitative results and conclusions.

We make a final observation on the way the quantitative/qualitative distinction has played out in the recent history of the discipline of geography. Like other social scientists, human geographers traditionally took a heterogeneous approach to methods before the late 20th century. However, as part of historical developments in geographic thought during the 20th century, polarized attitudes about proper methodology emerged within the discipline (see discussion and references in Chapter 1). Champions of the quantitative revolution and their predecessors pushed for a "scientific" approach, sometimes interpreted narrowly as relying solely on measuring observables quantitatively and precisely, and avoiding hypothetical constructs, subjective meanings, and other intellectual entities that they thought were too vague for a true positivist science. Various post-positivist critics of this approach during the latter half of the 20th century countered with their own somewhat extreme arguments that a strong positivist approach was not appropriate for understanding human experience, activity, and society. Among other things, this has produced rifts among geographers who focus on the study of

human activity and society, and especially between geographers who focus on humans versus those who study the natural earth (unfortunately, the word "versus" sometimes fits all too well).

We believe both of these positions are too extreme. Apparently unlike everything else in the world, humans and some other animals have **agency** (will)—to a degree, they determine when and how they act. Furthermore, because of brain evolution and cultural developments (including language and mathematics), human beings are, in part, **semantic** and **semiotic** entities—meaning and experience, often expressed in symbolic representations, partially guide their activities and explain their geography. Not all human geographers are required to incorporate such constructs in their work, by any means, but anyone who denies their relevance to geography is mistaken. We see no reason that scientific geographers have to ignore meaning and experience, although these constructs certainly create special intellectual and methodological challenges that biophysical scientists do not face. At the same time, there are unequal and unjust power relationships among subsets of people, cultural variations in conceptual structures, and idiosyncratic motivations among individual scientists for doing science. These do not, in our opinion, invalidate scientific approaches to understanding humans, although they certainly have implications for understanding how science should work and does work.

Review Questions

Primary and Secondary Data Sources

- What is the difference between primary and secondary data sources?
- How are primary and secondary data sources used in geographic research?

Types of Data Collection in Geography

- What are the following types of data collection in geography: physical measurement, observation of behavior, archives, explicit reports, computational modeling?

Quantitative and Qualitative Methods

- What are different factors that distinguish between quantitative and qualitative methods?
- What are some strengths and weaknesses of quantitative methods and of qualitative methods, and how do we recommend incorporating the two approaches into research?
- What role have ideas about quantitative and qualitative methods played in the 20th-century history of geography as an academic discipline?

Key Terms

agency: property of humans and some other animals of having at least partial self-determination of when and how to act

archives: type of data collection in which existing records that have been collected by others primarily for nonresearch purposes are analyzed, often after coding

behavior: the overt and potentially observable actions or activities of individuals or groups of people, or other animals

behavioral observation: type of data collection in which ongoing behaviors are recorded and analyzed, often after coding

coding: the process of categorizing qualitative records (such as behavioral recordings, archival records, and open-ended explicit reports) in order to turn them into analyzable data

computational modeling: type of data collection involving the output of a computational model, a model of theoretical structures and processes expressed in mathematical form, typically in a computer

explicit reports: type of data collection in which people's expression of their beliefs about themselves, other people, places, events, activities, or objects are recorded

physical measurement: type of data collection in which physical properties of the earth or its inhabitants are measured and analyzed

primary data: data collected specifically for the purpose of a researcher's particular study

qualitative methods: broad term referring to scientific methods that incorporate some combination of collecting nonnumerical data such as verbal or pictorial records, collecting data using relatively unstructured and open-ended approaches and formats, and analyzing data with nonnumerical and nonstatistical approaches; commonly used only in human geography

quantitative methods: broad term referring to scientific methods that incorporate some combination of collecting numerical data such as metric-level measurements, collecting data using relatively structured and closed-ended approaches and formats, and analyzing data with numerical and statistical approaches; commonly used in both physical and human geography

secondary data: data not collected specifically for the purpose of a researcher's particular study but for another research or nonresearch purpose

semantic: concerning meaning

semiotic: concerning entities or properties that represent or stand for other entities or properties, including signs, codes, symbols, models, and so on

study: unit of data collection and analysis activity focused on addressing a specific question or hypothesis

Bibliography

Abler, R. F., Marcus, M. G., & Olson, J. M. (1992). *Geography's inner worlds: Pervasive themes in contemporary American geography.* New Brunswick, NJ: Rutgers University Press.

Clifford, N., & Valentine, G. (Eds.) (2003). *Key methods in geography.* Thousand Oaks, CA: Sage.

Gaile, G. L., & Willmott, C. J. (Eds.) (2004). *Geography in America at the dawn of the 21st century.* New York and Oxford: Oxford University Press.

Haggett, P. (2001). *Geography: A global synthesis.* Harlow, UK: Prentice-Hall.