7

What is Data?

Chapter Summary

Previous chapters focused on how to find and formulate a research question (Chapter 4), how to develop a hunch about the answer to one (Chapter 5), and how to design research in order to investigate your hunch (Chapter 6). In the following chapters we discuss different methods of data collection and analysis. In this chapter, we focus on data. Our discussion focuses on the following issues:

- from concepts to measurement;
- validity;
- reliability;
- selection issues;
- types of data;
- big data;
- data quality checklist.

Introduction

If we want to develop our understanding about the political world or investigate different ideas or explanations about how things work, we need to think about how our ideas relate to what happens in practice. This means engaging with the empirical world—and to do this we need to collect information. Our record of the information that we have collected is data. We then use this data as evidence to support our arguments. Data thus constitutes the information base for our analysis. Without data—or evidence—argument is little more than conjecture.

In the following chapters, we will discuss different methods of data collection and analysis. However, before doing that, the purpose of this chapter is to outline what we mean by data in the first place. In this chapter, we discuss the main types of data used in Politics and International Relations (IR), and the main criteria by which we can judge whether our data is good, or not. Whatever type of data we use in our analysis, there are a number of issues we must confront. First, we need to ensure as far as possible that our data is valid. That is, the data does actually measure what we claim it does. And second, we need to ensure that our data is reliable. That is, we have recorded the data (or information) accurately. All observation and measurement in the social sciences is imprecise. We therefore need to take these issues very seriously. **Measurement error** can be thought of as the difference between the true value of our variable of interest and our recording of that value. Our data is never perfect—and so we must take steps to ensure that it is as good as possible. Good data helps us to make good **inferences** and draw strong conclusions. If the data is bad, then no matter how sophisticated our analysis, our conclusions will be found wanting.

From concepts to measurement

Research in political science usually starts with a hunch or a tentative theory about some aspect of political reality. To systematically investigate these hunches or ideas about different political phenomenon it is necessary to move from the theoretical or abstract world into the empirical world—and this means thinking about what data or evidence is relevant to answering your question (see Figure 7.1). There are a number of steps involved in this process. First, we must have a clear idea of what it is that we want to analyse. Suppose we are interested in analysing democracy—and whether a specific country is democratic or not (or has become more or less democratic over time). Democracy is obviously a very important concept, and one where there is considerable disagreement over what it means. Indeed, Michael Coppedge (2012: 11) refers to democracy as an *essentially contested* concept because although nearly everyone values the label, there are 'different reasonable and legitimate, yet incompatible, criteria for judging whether the label is deserved'.

To some people, democracy is simply a way of selecting a government; but to others it encompasses a much broader array of social and political rights. No definition of democracy will please everyone, but unless we want to get bogged down in semantic confusion we must nonetheless be clear about what we mean by the term. Before we move to data collection then, it is therefore important that we have a clear *conceptual definition* of what it is that we want to investigate, and what it is that we mean by the term 'democracy'. This process of deciding what we mean by a specific term is called conceptualization. Let's say that we want to adopt a minimalist definition of democracy and, following scholars like Vanhanen (1997, 2000), we define democracy as a system of electing governments through competitive elections, with full adult franchise. This definition emphasizes two central tenants of democracy to do with competition and participation.

The first challenge we face is to justify whether our conceptual definition is valid or not. There may be widespread disagreement about how to define a concept, so it is important to be clear at the outset what is included in our definition and what is excluded, and to provide

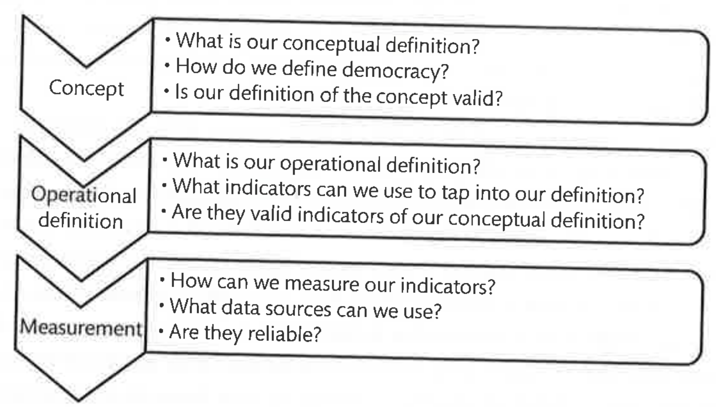

Figure 7.1 From concepts to measurement

a justification for the course of action we have taken. For example, we may acknowledge that there is more to democracy than simply holding elections, but that it is also true that democracy cannot exist without elections—and so, for the purposes of our research, we decide to focus on the minimum institutional conditions that must be met for democracy to exist. And, while not all countries that meet this minimum definition may be regarded as fully fledged democracies, clearly all those which fail to meet this condition can be regarded as non-democracies.

Having established a clear conceptual definition the next task is to think of an *operational definition*. Our conceptualization of democracy emphasizes competition and participation, and so we must think about how we can operationalize these terms. The process of operationalization involves moving from the abstract to the specific. Thus we must provide an operational definition of our two main components of democracy. For example, we may say that our operational definition of competition is the difference in vote share between the winning party in an election and the second placed party in that election. So, when one party wins by a huge margin we can say that elections are not very competitive, but when one party wins by just a few votes we can say that elections are competitive. For instance, Saddam Hussein used to regularly hold elections in Iraq, where he was reported to have received more than 99% of the vote. According to our definition of democracy, this would be an example of an 'uncompetitive election'.

Turning to our second indicator of democracy, our operational definition of participation might be the proportion of the adult population that turns out to vote. Thus, in countries where there is full adult franchise, and everyone is registered to vote, and then turns out to vote in an election, participation will be high. But in countries where some groups are not eligible to vote (such as women, or different ethnic or racial groups); or where there are barriers to registration (such as literacy tests); or where turnout is low; then participation will be low. For instance, during apartheid in South Africa only white people were allowed to vote and the black majority were barred from participating. According to our definition of democracy, this would be an example of restricted participation.

The main issue to consider with operational definitions is whether or not they are valid indicators of our concept. That is, once we have defined democracy as competition and participation, is the margin of victory in an election a valid indicator of competition? And is the level of adult participation in an election a valid indicator of participation? We may object to democracy being defined in such narrow terms, but concede that the operational indicators are valid. In this respect it is important to be internally consistent and to make sure that you do try and measure what you say you are going to study.

The last step in the process is to think about how we can actually measure our **indicators** of interest and what data we can collect. To measure competition, we may decide to simply inspect official election returns that are published after all the votes have been counted. Similarly, to measure participation we may count the number of votes that have been cast in the election and compare this to the most recent census estimates of adult population size in the country. The key issue here is to do with **reliability**. Are the election results an accurate reflection of what actually happened in the election? Maybe they have been manipulated in some way. Perhaps there has been electoral fraud. Similarly, does the country have reliable population estimates? Are there existing sources that you can use?

Validity

Generally speaking, whatever type of data is collected or used, the researcher will have to confront issues of **validity** and **reliability**. If the data itself is not valid or reliable, then the insights that are garnered from analysing the data are likely to be misleading in one way or another. Continuing with the example of democracy, the issue of validity is not whether we have measured participation and competition properly but whether these are suitable indicators of our concept—democracy. The validity of a measure therefore depends upon how we define the concept. This is always going to be open to debate, though is obviously more straightforward in some cases than others. There are no easy ways to resolve this issue. But there are a number of basic ways in which we can try and assess the validity of our indicators (see Moser and Kalton 1971, de Vaus 2002 for an extended discussion) and defend the approach we have taken.

The first and most straightforward is **face validity**. Face validity simply means that, on the face of it, the indicator intuitively seems like a good measure of the concept. Thus we may all agree that participation is an important element of democracy and so has a certain amount of face validity as an indicator; but if we were to suggest that we could measure democracy by looking at incarceration rates we might question what, if anything, this has to do with our central concept of interest. Face validity thus refers to whether there is broad agreement that the indicator is directly relevant to the concept. If the answer is 'no' then we definitely have a problem, but what is intuitively 'yes' for one person may not be so for another.

Content validity examines the extent to which the indicator covers the full range of the concept, covering each of its different aspects. Thus, we might argue that the indicators of democracy lack content validity as democracy encompasses more than just participation and competition. Our measure therefore neglects to take into account important political rights, such as the right to self-expression, the right to seek out alternative forms of information, and the right to form independent associations and organizations which scholars such as Robert Dahl (1989) have suggested are important preconditions for the presence of democracy. When only one or two dimensions of a multifaceted concept are measured, it is important to be explicit about which dimensions have been left out (and why), as well as which dimensions have been included. However, in general it is good practice to avoid overclaiming what your indicators measure. So if you are just measuring political competition, it is better to be explicit about this at the beginning—and say that your project is concerned with analysing why some countries are more competitive than others, rather than why some countries are more democratic than others.

A third way to evaluate the validity of an indicator is by empirical analysis. **Construct validity** examines how well the measure conforms to our theoretical expectations by examining the extent to which it is associated with other theoretically relevant factors. So, on theoretical grounds we could assert that democracy should be associated with freedom of the press. Thus, we might not regard a free press as being part of our definition of democracy, but nonetheless expect democracies to respect press freedoms more than non-democracies. And so, in order to validate our measure of democracy we would see whether it is associated with press freedom. The presence of such a relationship lends some support to the validity of the measure. Ultimately, though, the matter of validity is one of judgement.

When we deal with difficult-to-define or multifaceted concepts, these issues are difficult to deal with. It is therefore sometimes tempting to skirt around them and avoid making precise statements about what we are doing, or why. However, this approach leads to very vague and impressionistic conclusions, which lack a strong empirical core. When a concept is difficult to define it is more important than ever to be explicit about how we are going to operationalize it. It is only when we have a clear idea of what it is we want to find out that we can begin to think about what sort of data might be useful. Even then, it is sometimes very difficult to identify what type of data we should use, and whether there is existing data we can analyse or new data we can collect.

Example: measuring corruption

We can illustrate some of these issues with recent attempts to measure corruption. Corruption is a major topic of interest in Politics and International Relations (IR), but it is also a difficult topic to analyse because, by definition, corruption is a clandestine activity that is difficult to directly observe. Any study that wishes to investigate corruption therefore has to confront a series of difficult questions about how to operationalize and measure the incidence of corruption in different parts of the world (for a recent review, see Heath, Richards, and de Graaf, 2016).

The first step is to be explicit about what is meant by corruption. A simple and widely-used definition, used by Transparency International, the World Bank, and the US Department of Justice, is that corruption is 'the misuse of public position for private gain'. Typically, this involves paying a government official a bribe in order to receive goods or a service. The problem is—because of the clandestine nature of the activity—it is very challenging to think how we could actually measure this type of activity. Obviously we could just ask public officials whether they ever demand bribes, but it is doubtful whether they would reply honestly. Instead, researchers have concentrated on trying to develop a number of indirect measures of corruption (see Figure 7.2).

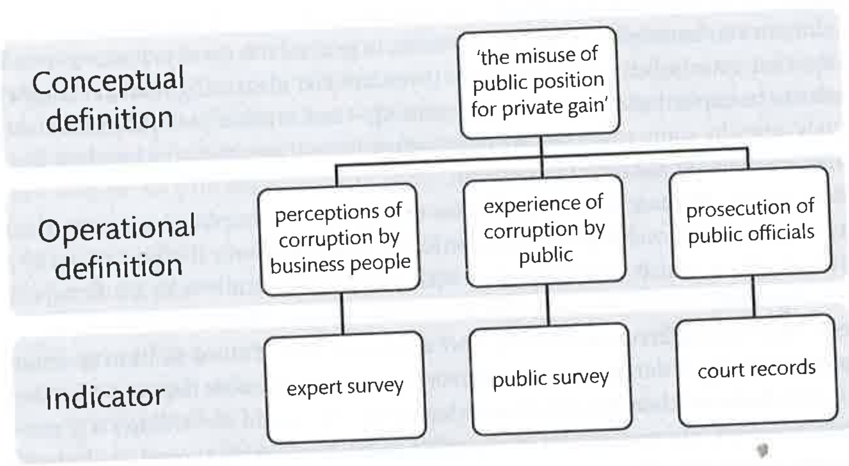

Figure 7.2 Approaches to measuring corruption

One widely used approach is to rank countries according to their level of 'perceived' corruption. For example, Transparency International regularly carries out a survey which asks businessmen to rate the level of corruption in different countries. Thus, the operational definition of corruption is not strictly the incidence of corruption but the perception of corruption among a group of people who should be well placed to judge the business practices of different countries around the world. This kind of 'perception' data is widely used in Politics and IR. However, a major issue with this type of data is whether perceptions of corruption are a valid indicator of the actual incidence of corruption. Does the perception that a country is corrupt mean that a country really is corrupt? People's perception of the level of corruption in a country might be shaped by all sorts of factors—such as their political ideology, exposure to media, or even whether they like the country or not and enjoy spending time there—so the indicator may not be particularly valid.

So, what other types of data could therefore provide a better source of information? Some scholars have used conviction statistics to measure how many public officials are successfully prosecuted on charges of corruption (e.g. Glaeser and Saks 2006 on Federal prosecutions in the USA). But this indicator is also problematic as they may just reflect the energy and commitment of law enforcement agencies rather than the actual incidence of corruption (Goel and Nelson 2011). Lastly, other scholars have used 'victim' surveys to measure whether or not members of the public have actually had to pay a bribe in order to receive a good or service. This issue of validity remains with victim data, though in many respects they are the least problematic of the measures currently in use (Heath, Richards, and de Graaf 2016).

Reliability

Related to the issue of validity is the issue of **reliability**. Broadly speaking, reliability refers to how accurately we have measured our indicator. Thus, even if we have a valid indicator, if we are not able to measure it accurately we may end up with unreliable data. In more technical terms reliability means the measurement of an indicator yields the same results on repeated occasions. The higher the reliability of the measuring procedure the more consistent the results will be. Data can be valid but not reliable, and can be reliable but not valid. Obviously we would prefer it if our data was both valid and reliable.

There are a number of different ways of assessing the reliability of our data—and these techniques vary somewhat according to the type of data we are interested in. To illustrate some of the more common ways of doing this let's continue with the example of measuring corruption. We may have concerns about the validity of using perceptions of corruption to measure actual corruption; but should we also have concerns about the reliability of this approach? If the data is reliable then there should be broad agreement between the businessmen about which countries are corrupt and which countries are not. If there is a lot of disagreement between businessmen about a country's level of corruption—and some people, for instance, regard business practice in India as very corrupt, whereas others do not—then our data may be unreliable. We refer to this type of reliability as **intercoder reliability** which is discussed in more detail in Chapter 14. Intercoder reliability reveals the extent to which different coders (in this case businessmen), each coding the same content (a country's level of corruption), come to the same coding decisions (whether the country is corrupt or not).

Another way in which researchers often talk about reliability is whether our measurement of our indicators produces consistent results on repeated occasions. This is somewhat different to intercoder reliability and is generally used in different contexts. So, for example, if we wanted to know whether the 'victim' surveys produced reliable data on whether people had experienced paying a bribe to a government official we might want to ask our respondents questions about this on more than one occasion. This is sometimes known as the *test-retest* method. It involves applying the same test to the same observations at different moments in time and comparing the results. We discuss this in more detail in Chapter 11. If all the tests produce the same result we have a reliable measure—but if they do not then we don't. It is clear that unreliable data lacks validity. But reliable data need not be valid, because it could be measuring something other than what it was designed to measure, and high reliability should not compensate for low validity.

Selection issues

In addition to these issues, all data need to confront the issue of *external validity*. This means that we have to think about how far we can generalize the findings from our data. Sometimes people will make grand generalizations, say that they have discovered some universal truth, and that their findings apply across time and space. However, more often we will make bounded generalizations and say that our findings apply to a particular country at a particular point in time; or perhaps even to just a particular group within that country. Making generalizations doesn't mean that your findings apply to everyone (or everything) but that your findings apply to more than just the people you have spoken to or the particular cases you have collected data on. If we are unable to make any generalizations—or wider inferences—then our data is little more than an anecdote.

There are often thought to be two trade-offs with respect to data collection. On the one hand we want our data to be **internally valid** and to measure what we are interested in. So, to take another example, if we want to know what someone thinks about immigration, we might think it important to have a detailed conversation with them; talk to them about their experiences, perceptions, and perhaps even observe their interactions. But, we often want to draw conclusions as well, and in this respect it is important that our data is **externally valid**. While it is interesting to know what our neighbour thinks about immigrants, we might also wonder how typical our neighbour is and whether their views are also shared by others. This brings us to selection effects and how far we can make generalizations from the data we have collected.

In order to set the parameters for generalization, it is important to think about case selection. This is a critical, though often underappreciated part of data collection, and to a large extent determines what is good data (or at least useful data) from bad data. We have already discussed the issue of internal (or measurement) validity: our empirical indicators must be closely related to our concept of interest. In a similar way external (or selection validity) means that the cases we select are related to the population from which they are drawn. When we collect data we often deal with samples, however big or small those samples are. For example, if we are doing a study on young people, we do not interview all the young people in the world, or even in one particular country. Rather, we select a sample to interview.

This sample might be a few thousand people if we are conducting a survey—it might be just a handful of people if we are conducting a semi-structured interview. The people that we end up interviewing represent our cases, and for each case we collect data.

The issue of case selection is extremely important because the cases we choose to analyse and collect data on can influence the answers we get to a particular question. When this occurs, it is known as **selection bias**. If we do not think carefully about case selection we can end up with answers that are not very robust, or are even misleading. That is, we will make faulty inferences. The best way to avoid this happening is to conduct a large-N study and select cases using random probability methods, or to take a full census of cases (see Chapter 11). But this is not always a viable option. To a certain extent then, selection bias, or at least the risk of selection bias, is always going to be a danger in small-N studies, particularly when the study is designed to try and make some sort of generalization about a wider population, however tentatively put. We discuss the issue in more detail in Chapter 9 and Chapter 11.

In order to minimize these threats to external validity it is important to follow a number of steps (see Figure 7.3). Suppose you were interested in examining when the UN Security Council intervenes in domestic conflicts. You might have a range of hypotheses about what motivates the UN to intervene in some conflicts but not others, from humanitarian need to political expediency (see Binder 2015). To investigate your hypotheses you will need to choose some cases to look at and collect data on; but which cases will you look at and how will you choose them? Sometimes students will rush to focus on one or two high-profile cases without considering the broader range of cases they could potentially study. The first step then is to define your population of interest. The population in this context refers to all the possible cases you could theoretically pick to analyse. Since a case is a domestic conflict, the population for your study would be all the domestic conflicts that have taken place in the world since, say, 1946. This is obviously going to involve a lot of cases, and it would be a lot of work to find out how many conflicts there have been from scratch. The Peace Research Institute Oslo (PRIO) keeps a detailed log of all such conflicts.

The next step is to outline a clear rationale for deciding which particular case or cases you are going to look at. The key point is that there must be some rationale for selecting some cases, but not others—and this rationale should be related to the theories or ideas that you want to investigate. There are various ways of doing this which we discuss in more detail in Chapter 9—but unless there is a strong justification for case selection it is very difficult to know what the broader implications of your research are likely to be. For instance, you could

Figure 7.3 From populations to cases

decide to pick a few cases where there has been a clear humanitarian crisis, and use this as a basis for investigating whether political considerations enter in the calculus of whether or not the UN decide to intervene. Having decided upon selection criteria, the last step is to pick the cases that you will actually investigate. Following this approach allows you to contextualize your data and draw stronger conclusions.

Although this approach to case selection sounds quantitative, the way in which we choose a sample depends more upon the type of information we want to find out and the nature of the conclusion we want to draw than it does on the type of method that we use. Large samples are not always necessary, or even desirable, and valid conclusions can still be drawn from analysing just a few cases, but only for particular types of research questions. For example, suppose you wanted to find out about the policymaking process in a particular country. You could approach this topic in a number of different ways depending upon what you wanted to find out. First, suppose you wanted to find out how the policymaking process works. To do this, you might want to interview 'experts' who have first-hand knowledge about the policymaking process (e.g. policy-makers themselves). You could just interview one policy-maker and ask them how the process works. Hopefully, the expert you have chosen to interview is competent and can give you a reasonably accurate account of the process. However, just to be on the safe side you might want to interview more than one policy-maker. If you interview five policy-makers and they all give you a similar account about how the policymaking process works, you can be reasonably confident in the data you have collected. Thus, you can make a valid **inference**—or a valid conclusion about your object of interest (the policymaking process) from just a handful of interviews. But suppose you had a slightly different research objective and wanted to know what policy-makers *thought* about the policymaking process and whether they thought it worked well or not. Now you would want to make a very different sort of inference—not about how something works, but about what people think, and this means you would want to draw a conclusion about a broader population of interest (policy-makers). In this situation you would be very unwise to draw strong conclusions about what policy-makers as a whole think about the policymaking process based on interviews with just five people.

Types of data

The form that data (or information) takes can vary, from surveys to interview and focus group transcripts, ethnographic field notes, historical records and speeches, to government statistics, to name but just a few. Political scientists, perhaps more than researchers in any other discipline, draw on a wide and varied range of sources of information to study the political world. And just as the types of data used vary, the techniques used to analyse the data also vary. In this section we discuss the main types of data that are used in Politics and International Relations and provide a brief overview of the issues they raise with respect to validity, reliability, and selection bias.

It is common to distinguish between primary and secondary data. *Primary data* is often associated with data we collect ourselves. That is, data collected, recorded and analysed by the researcher carrying out the project. By contrast, *secondary data* is often associated with data collected by other people. That is, the researcher analyses (or re-analyses) data that has been collected by someone else. The distinction between these two types of data relates more to the issue of who collected the data than it does to the type of data in question. For example, we could conduct a survey ourselves, in which case the survey would be an example of primary data. Or we could analyse a survey that has been conducted by someone else (such as the British Election Study), in which case the survey would be an example of secondary data. In both cases we are dealing with the same type of data, which has the same strengths and weaknesses.

The main advantage of collecting primary data yourself is that you have full control over data collection, and so are aware about any issues that might affect data quality (such as a low response rate to your survey). The main disadvantage is that data collection can sometimes be a time-consuming and costly exercise—and you may not have the resources to collect the data that you need. However, often we can use secondary data to answer the questions we are interested in. The UK Data Archive is an incredibly valuable resource, and contains records of thousands of studies (both quantitative and qualitative) on a wide variety of different topics that are freely and publicly available. The main disadvantage of using secondary data—or at least the main issue to be aware of—is that it is not always easy to tell whether the data is of high quality. It is therefore important to only ever use data you have confidence in, where there are clear methodological statements about how the data was collected, who it was collected by, and importantly, who it was funded by and why it was collected.

Data can be categorized in many different ways. If we want to find out information about something we have three main options. We can ask someone a question, we can observe what they do, or we can read what they have written. These different strategies form the main methods of primary data collection that we use in political science: interviews; observation; and document analysis—and we discuss each method in-depth in the following chapters. Within these broad categories of data collection there are many distinct approaches and we can distinguish between those that produce quantitative data and those which produce qualitative data (see Box 7.1). In addition, researchers frequently use *official statistics* or data that has been compiled by government agencies or international organizations (such as the World Bank). This type of secondary data raises a number of distinctive issues, which we discuss later.

BOX 7.1 Different Types of Data

Method of collection	Type of data	
	Quantitative	**Qualitative**
Interview	Results from a survey	Transcripts from interview or focus group
Observation	Roll call votes from parliament	Field notes from participant observation
Documents	Policy position of parties, salience of issues	Party manifestos, speeches, policy documents, diaries
Secondary sources	Government statistics	Government reports

Quantitative and qualitative data

One of the most common ways of classifying data is to distinguish between quantitative data, which is numerical, and qualitative data, which is not. Whereas quantitative researchers assign numerical values to their data, qualitative researchers tend to assign words (King, Keohane, and Verba 1994: 151). Thus, even when researchers are looking at the same source—for example, a party manifesto—they may choose to represent the data in different ways (see Chapter 14). The primary differences between qualitative and quantitative data thus relate to how the data are represented (or coded) rather than what type or source of data is necessarily being analysed.

Sometimes people get very animated about whether or not qualitative data is better than quantitative data. Some people might think that numbers tell you nothing, and so the only way you can ever find something out for sure is to do a detailed qualitative study and describe what you see using words. Others think that qualitative studies prove nothing, and so like to just rely on numbers. There are not many areas of life where we draw such firm distinctions between different types of evidence—and wilfully ignore what we don't like the sound of. Most of the time, when we care about getting the 'right' answer to a question we will consider a variety of different types of evidence—or data. Suppose we are buying a house, we might consider quantitative data (how much it costs; how big the house is) and qualitative data (what the house looks like, whether it is light and airy, or has a 'wow factor'). It would be a strange person that decided to buy a house without knowing how much it cost or without seeing what it looked like. Similarly, when we investigate a political phenomenon it makes sense to consider all types of relevant information regardless of whether it is qualitative or quantitative. Most examples of good research in Politics and IR draw on multiple sources of qualitative and quantitative data. And even when the data presented is explicitly quantitative, such as the results from a large-scale social survey, often a great deal of qualitative research has gone into the process of developing and testing the questions during various rounds of pilot studies.

That said, there are some important differences between qualitative and quantitative data. In very blunt terms, these differences involve a trade-off between internal validity and external validity. Qualitative researchers try to achieve accurate measures, but they generally have somewhat less precision (King, Keohane, and Verba 1994: 151). Thus qualitative research is a very good way of assessing and establishing the validity of the data that we use—even if that data is quantitative in form. But since qualitative data tends to prioritize detail over theoretical generality—it is often very difficult to tell what the broader implications of the data are—and whether similar data on different cases would produce consistent results. Part of this is simply just a small-N problem (see Chapter 9). This can either be overcome gradually by **replication**, where similar data is collected in different contexts to see if the findings can be applied in the same way. Or it can be overcome by deriving testable hypotheses from the qualitative data that can then be examined more generally using quantitative data (see Chapter 13). In either case, qualitative and quantitative data can often be combined to good effect—and so the distinction between the two types of data is perhaps not as divisive as is often portrayed. Data also varies with respect to other characteristics, and focusing purely on whether the data is qualitative or quantitative risks downplaying the salience of these other differences.

Interview data

Interview data is perhaps one of the most widely used forms of data in political science. Much of our research concerns what people (or actors) do, say, or think. One particularly useful way of finding out this information is to ask questions to those concerned. There are many different ways to do this. Interview data can be gathered using more, or less, structured methods: for instance, using open-ended and flexible questions in semi-structured interviews or focus groups (Chapter12), or questions to be answered by respondents selecting from among a fixed set of choices in surveys (Chapter11). The main instrument for gathering data might be the researcher herself, or it may be that data are not 'collected' at all, but rather are co-produced by the researcher and the subject as part of an ethnography (Chapter 13). Interviews might be carried out with a representative cross-section of a given population, or with experts who have specialized knowledge about a given topic.

The actual data that these methods of collection produce are the specific responses to the questions that respondents provide. These responses might be recorded in different ways: they could be audio recordings on digital devices; they may be transcribed verbatim in written documents; or they may be coded using numbers in databases. For example, a researcher may be interested in the opinions of young people towards electoral reform. They could try and find out about these opinions by collecting data through surveys, or through interviews, or focus groups, or by ethnography. In each case the researcher could ask the same question. For example: 'what do you think about the House of Lords? Do you think it needs reform?' This is obviously a very general open-ended question which people might interpret in a number of ways and give a variety of different answers to. Some people might say that they don't know or that they've never thought about it, or they have no idea. We could write down each of these responses word for word. We might think that this sort of data is qualitative; but if we decide to code each of the answers into broad response categories, such as 'no opinion', we can easily ascribe a numerical value to each category and transform the qualitative data into quantitative data. Different types of interview data can therefore be integrated together to give a fuller picture of what young people actually think.

Observational data

Whereas interview data is based on the ancient art of asking questions to find out what people think, say, and do, observational data is based on something rather different. It recognizes that what people say they do, and what they actually do are often different. For example, someone may claim that they are not sexist, but nonetheless may frequently behave in a sexist way. Accordingly, to get a 'true' sense of what people think and say and do, it is not enough to merely ask people questions and record their answers; it is also necessary to observe what people do in practice.

There are various ways in which political scientists do this. One form of observational data comes from participant observation (Chapter 13). For example, in a classic political ethnography Richard Fenno (1978) conducted research in the USA with eighteen members of Congress as they travelled around their constituencies, spending time with them and observing what they did and who they spoke to on a day-to-day basis, providing fascinating insights in to how Congressmen interact with their constituents, and how this translates into their legislative effectiveness. Another form of observational data comes from collecting systematic data on the behaviour or actions of different actors. This data may be collected by researchers or recorded by different organizations. For example, Keith Poole and Howard Rosenthal (2000) analysed 16 million individual roll call votes in the Houses of Congress since 1789 to examine the voting patterns of legislators in the USA. Another form of observational data comes from experiments—when participants are exposed to different stimuli to see how they react (see Chapter 8). Different types of observational data can therefore be integrated together to give a fuller picture of how people actually behave.

Documents

With interviews and observation the data is collected—or generated—by the researcher. That is, through asking questions and recording answers, or by observing and recording patterns of behaviour, the researcher is able to produce new data that can be analysed in order to answer the key research questions. Other types of data do not need to be collected in the same way. For example, researchers may analyse written policy documents or political speeches, media reports or broadcasts, personal diaries—or, more recently, social media posts from both a qualitative and a quantitative perspective (see Chapter 14). Analysing documents is generally an unobtrusive method of data collection (though this is not quite so clear-cut when it comes to social media posts). In this case the raw data is the document itself. Sometimes researchers will analyse the documents from a qualitative perspective, and extract excerpts or quotations to illustrate a point. Other times researchers will analyse the documents from a quantitative perspective and code the content of the document into distinct categories to produce new data.

For example, the Manifesto Project provides the research community with parties' policy positions derived from a content analysis of parties' electoral manifestos (Volkens et al. 2015). It covers over 1,000 parties from 1945 until today in over fifty countries on five continents. In each manifesto, the policy statements in each ('quasi-') sentence are classified into fifty-six policy categories over seven policy domains. The coded data provide a useful indication of party positions since they represent the choices that the electorate faces before each election. These measures are generally consistent with those from other party positioning studies, such as those based on expert placements, citizen perceptions of parties' positions, and parliamentary voting analyses, which provides additional confidence in the validity and reliability of these estimates (Laver, Benoit, and Garry 2003).

Secondary data

Whereas the types of data discussed earlier can be primary or secondary, other types of data are purely secondary. Political scientists frequently use data or sources that they have not collected themselves, and which have been collected or produced for purposes other than academic research. This type of data often relies upon the record-keeping activities of government agencies, interest groups, and international organizations (see Johnson and Reynolds 2008: Chapter 9). For example, various government agencies routinely publish official statistics which document the level of poverty, unemployment, and crime within the country. Although these government statistics can be useful, they also need to be treated with

caution. As with all secondary data, the researcher is not directly involved in the process of data collection, and so is in a sense at the mercy of the practices and procedures that have been adopted by the organization that has collected the data. These practices and procedures are not always transparent. One particular issue to be aware of is bias. Government statistics on issues such as crime or poverty can be very sensitive—and in some cases government agencies may try to 'massage' the figures to present the government in a more favourable light. This can make comparative research very difficult—as governments in some countries may simply be more willing to publish accurate data than others. Moreover, changes in the way crime or poverty are classified can affect the level which is recorded. This can make analysis difficult over time. As always then, when using secondary data it is important to assess its validity and reliability and to be aware about potential biases.

Big data

So far we have considered the main types of data that are used in Politics and International Relations. In recent years however, there has been a much heralded 'revolution' in the way that data is collected and analysed. This so-called 'big data' revolution challenges many of our assumptions about what data is, how data is generated, and what data can tell us about the social and political world. In this section we discuss what big data is, how it is being used, and its potential applications for research in Politics and IR. We also examine some of the problems and pitfalls with big data, and discuss the differences between big data and other types of data.

Big data is a very modern phenomenon. Although political scientists have long worked with very large datasets—such as Census data which contains data on millions of people—it is only relatively recently that people have started to speak about 'big data' as a separate category of data. As Burt Monroes (2012) explains, as political scientists we are used to thinking about 'small-N' or qualitative studies and 'large-N' quantitative studies, and so it is perhaps natural to equate big data with 'really large-N' studies and to think about it purely in terms of scale and size. However, this misunderstands the 'big' in big data and the qualitative difference that exists between 'big' data and other types of data. At the turn of the century, IT analyst Doug Laney laid out a framework for understanding the challenges that were then emerging in data management for business, now widely referred to as 'The Three Vs of Big Data': Volume, Velocity, and Variety (Laney 2001). Simply put, data were beginning to be created at a scale, speed, and diversity of forms not seen before. According to Gary King the 'big data revolution,' is a technological revolution that has transformed our ability to produce, collect, store, and analyse vast amounts of data. One of the key features of big data is that the data is generated almost spontaneously. Vast amounts of activity now leave a digital footprint, and these digital records constitute a sort of data that can be readily harvested. Researchers do not therefore need to generate or collect the data in the traditional sense. For this reason big data is sometimes described as 'found' data. The data is already there, and all the researcher needs to do is harvest it. This also makes big data incredibly cheap compared to similar more traditional forms of data. Traditional face-to-face surveys are very expensive. The British Election Study for the 2015 elections cost over £1 million. By contrast, the opinions of millions of people on Twitter are free to harvest.

As Gary King (2013) has said, big data offers the opportunity to study things we have always looked at from a new perspective and to study things we previously found difficult to analyse (see Box 7.2). Perhaps one of the most widespread uses of big data is to mine social media. Traditionally, if we were interested in the political opinions of people we would carry out a survey—of say 1,000 people—and ask them what they thought about a particular issue. This form of data is not only relatively expensive to collect, but also time-consuming. Now, with social media, millions of people readily volunteer their opinions on matters of the day. These opinions can be freely harvested from platforms such as Twitter in a matter of seconds. Scholars such as Ben O'Loughlin have used this approach to investigate how people on Twitter respond in real-time to high-profile political events. One such event he looked at was the controversial appearance of Nick Griffin, the former leader of the far-right British National Party, on the weekly British political debate show BBC Question Time (O'Loughlin and Anstead 2011). During the course of the TV show, some 50,000 tweets were posted about the programme which were then harvested and coded, revealing which segments of the programme produced the biggest 'spikes' in Twitter activity.

Other examples of big data involve harnessing mobile technology. Traditionally, if we wanted to know how much exercise people took the only way we could find out was by carrying out a survey and asking people what forms of activity they undertook. Now, researchers are using mobile phones with special apps to actually record how much hundreds of thousands of people move around each day, measuring the distance they travel and the speed at which they move using accelerometers. Indeed, there are many innovative ways in which big data is being used and the digitalization of vast catalogues of archival materials is blurring the boundary between qualitative and quantitative research and forging new collaborations between historians and computer scientists. For example, one ingenious study was carried out by Dan Cohen and colleagues on trial records from the Old Bailey from the seventeenth century as part of the Digging into Data programme. The Old Bailey website (http://www.oldbaileyonline.org) makes available 127 million words of accurately transcribed historical text, recording the details of 197,000 trials held at the Old Bailey or Central Criminal Court in London, between 1674 and 1913 (Hitchcock and Shoemaker 2006). All manner of people,

BOX 7.2: Uses of Big Data (King 2013)

Topic	Traditional data	Big data
Opinions of political activists	Survey—how interested are you in politics?	Analysis of Twitter: billions of political opinions in social media posts
Exercise	Survey—how many times did you exercise last week?	500K people carrying cell phones with accelerometers
Social contacts	Survey: Please tell me your five best friends	Continuous record of phone calls, emails, text messages, social media connections
Economic development in developing countries	Dubious government statistics	Satellite images of human-generated light at night, road networks, other infrastructure

across all ages, genders, and classes appeared in London's court. Their words, collected in the evidentiary statements (comprising 85% of the proceedings), detail crimes, their ancillary activities, their perpetrators, and their victims. Dan Cohen, for instance, looked at cases that involved women in the nineteenth century and was able to chart changing attitudes towards gender by examining how women were punished for different offences. By looking at cases at the Old Bailey involving bigamy he could see a significant rise in women taking on other spouses when their husbands had left them. It also became very clear that these late-Victorian women were not receiving significant punishment for this action, as they did earlier in the century.

Although big data opens up many new avenues for research, it also presents some familiar problems to do with reliability, validity, and bias. Just because there is more data, it does not mean that these problems go away, and indeed, they can even be amplified. Perhaps one of the biggest problems with big data is to do with selection effects. Big data is often seen in a favourable light by policy-makers because it provides a low-cost way of gathering information which can then be used to allocate resources. However, failing to take selection issues into account can lead to serious problems, and inequitable outcomes. A good example of this comes from Boston in the USA which had a big problem with potholes in the city streets, with 20,000 potholes needing to be filled every year. This put a considerable drain on the city finances, not least in terms of identifying where the potholes were in the first place. To help allocate its resources efficiently, the City of Boston released the Street Bump Smartphone app, which draws on accelerometer and GPS data to help detect potholes, instantly reporting them to the city. This was an ingenious use of big data technology. However, there was a problem. People in lower income groups in the US are less likely to have Smartphones, so the people using the app tended to be well off, and consequently the potholes identified tended to be in well-off areas or on popular commuter routes. This meant that the potholes in poorer parts of the town were less likely to be noticed and therefore less likely to be fixed—hardly an equitable approach. Fortunately this problem was soon identified and so measures were put in place to try and redress the digital divide.

Other examples of big data failing to live up to its initial promise are illustrated by the now famous parable of Google Flu trends (Lazer et al. 2014). Outbreaks of flu—or mini epidemics—are common in many countries, and these can put a strain on medical services. Identifying when and where outbreaks are likely to occur can therefore be an enormous help for distributing resources to where they are most needed. Traditionally local doctors would report cases as they came in—but this process tended to be slow and so did not provide much of an early warning signal. Google realized that outbreaks of flu could be predicted with a remarkable degree of accuracy by looking at the search terms that people entered into Google. But the problem was that this was a largely inductive approach. Google identified the search terms that best predicted outbreaks of flu, whether those terms were directly related to the symptoms of flu or not. There was thus little theory driving the data collection. Although the algorithm they developed was highly successful at first, it soon stopped working, serving as a timely reminder that correlation does not necessarily equate to causation.

The approach that we have described in this book is theory-led. We come up with an idea—we engage with theory to form hypotheses—which we then try and operationalize in order to test. Some commentators have suggested that big data can replace this scientific method, with theory and **research design** essentially becoming obsolete. The editor-in-chief of *Wired*

Magazine, for example, writes that 'Correlation is enough. …We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.' (Anderson 2008). However, this approach is almost certainly doomed to failure over the long run. Political scientists have typically had more reservations about the promise of big data than their computer scientist counterparts, largely because of their training in statistical theory, which makes them aware of issues related to things like sampling populations, confounders, overfitting, and multiple hypothesis testing (Clark and Golder 2015). As Manski (2013) and Keele (2015) observe, conclusions result from a combination of data and theoretical assumptions—without assumptions, data is just data. As Clark and Golder (2015) put it, the choice confronting the researcher is therefore not between induction or deduction, or description and explanation, but between observation guided by conscious theorizing and observation guided by unconscious theorizing.

Data quality checklist

Whatever type of data is collected or used there are a number of fundamental principles worth addressing to convince yourself and your reader that the data is of good quality and that the findings can therefore be trusted (King, Keohane, and Verba 1994: 23). The process of establishing the validity and reliability of our data is important for determining whether or not we can have confidence in our results. Because we use the data to test our hypotheses we must have confidence that the data is as accurate as possible. If our data is not accurate then we may come to faulty conclusions, and either incorrectly reject a hypothesis when it is in fact true, or incorrectly accept a hypothesis when it is in fact false. Obviously we want to have confidence in our own research, but we also want others to have confidence in it too and to take the findings seriously.

The first and most important guideline is to be transparent about how the data has been collected and to report and record in as much detail as possible the processes by which the data have been generated. One way in which we can give others confidence in what we have done is to be as open and transparent as possible about what we have done at each stage of the data collection and analysis process. In this respect, the principles of replication and verification are very important. It is now a widely accepted practice to make all publicly-funded data publicly available. This is now a funding requirement in many countries around the world. Public bodies fund research and so demand that the data be shared amongst the community. This means that the data collected by individual researchers or teams of researchers is freely available to other scholars who may wish to analyse it. Although quantitative research has led the way in making data publicly available, qualitative researchers are now also being encouraged to be more transparent about their data collection. Datasets and interview or focus group transcripts with detailed methodological appendices are now routinely uploaded to project websites and are deposited in digital archives such as the UK data archive.

Making data available in this way creates a public resource that other scholars can share. It also makes the research process transparent. **Verification** refers to the practice of making data available for secondary analysis so other scholars can re-analyse the data to see if they come to the same conclusions. **Replication** refers to the practice of carrying out a similar data collection exercise to see if the findings from the original study can be reproduced on an independent data set. Replication—whether it is for qualitative or quantitative data—is the primary way in which we can establish the external validity of our data.

These principles of replication and verification make the research process more transparent, which in turn gives us confidence in the results as it adds an extra layer of scrutiny to how the data has been collected and analysed. This openness and transparency allows us to develop knowledge by building on each other's work. It also allows us to identify cases of misconduct. This scrutiny has helped to uncover some major instances of research fraud. In a now famous case, in 2014, an article was published in *Science* by PhD student Michael LaCour and Professor of Political Science Donald Green, which claimed to show that personal contact with gay rights activists had long-lasting effects on people's attitudes towards gay marriage. In the process of trying to replicate the study—and re-analyse the data that had allegedly been collected, another PhD student, David Broockman, noticed some irregularities which led him to suspect that the data had been fabricated. These suspicions were corroborated when the survey agency who had apparently collected the data were contacted, said that they knew nothing about the study, and that they had never worked with Michael LaCour (who was the principal investigator in the study). The article was later retracted from the journal. This scandal undoubtedly did a lot of damage to the credibility of the discipline. However, the fact that it was uncovered shows the value of replication and verification. Although replication and verification cannot ensure that irregularities do not take place, they can help to uncover them. And without replication and verification it is highly likely that more fraud would occur. As in other aspects of political life, transparency is the best antidote to corruption.

Conclusion

Politics and International Relations is about making arguments and supporting those arguments with evidence. The data that we collect provides the information base for these arguments. If the evidence is bad then the argument is misleading. In order to have confidence in our results it is important to ensure that our data is as good as possible. And so that others may also have confidence in what we have done, it is important to be as transparent and open as possible about how our data have been collected and how they have been coded. Although there are a tremendous amount of different types of data that we can draw on, the principles of what makes good data are similar. Whether we are interested in qualitative interviews or big data analytics, the issues of selection, measurement, theory, and validity are equally important.

Questions

1. How can we measure difficult-to-define concepts in Politics and International Relations?
2. Why is measurement important?
3. Is there a trade-off between internal validity and external validity? Which should we be most concerned about?

4. What is the difference between qualitative and quantitative data? How important are these differences?
5. Is it always better to collect data on a large number of cases? When might data from just a few cases lead to valid conclusions?
6. How would you assess the validity and reliability of secondary data? What questions would you want to ask?
7. Is big data different to other types of data? What, if anything, is distinctive about it?
8. Should data be made publicly available? What are the potential drawbacks to making data available?

Guide to Further Reading

Clark, W. and M. Golder (2015), 'Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?', *PS: Political Science and Politics* 48(1): 65–70.

This special issue contains a number of articles which consider some of the problems and pitfalls of big data and the challenges it presents to conventional methodological approaches in political science.

Coppedge, M. (2012), *Democratization and Research Methods* (New York: Cambridge University Press).

Provides an in-depth but accessible discussion about the problems of operationalizing and measuring democracy, and the consequences for analysis.

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).

Although this book is primarily about surveys, it contains very clear discussion about measurement issues, operationalization, and the principles of validity and reliability, which are of relevance to other methods of data collection.

Heath, A., L. Richards , and N. D. de Graaf (2016), 'The sociology of corruption', *Annual Review of Sociology* 42.

Provides a detailed discussion about the strengths and weaknesses of different approaches to measuring corruption, and outlines promising avenues for further research.

King, G., R. Keohane , and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).

This classic text provides a thorough examination of measurement and selection issues in qualitative and quantitative research.

References

Anderson, C. (2008). 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired Magazine* 16.07.

Binder, M. (2015), 'Paths to Intervention: What explains the UN's Selective Response to Humanitarian Crises?', *Journal of Peace Research* 52(6): 712–26.

Clark, W. and M. Golder (2015), 'Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?', *PS: Political Science and Politics* 48(1): 65–70.

Coppedge, M. (2012), *Democratization and Research Methods* (New York. Cambridge University Press).

Dahl, R. (1989), *Democracy and Its Critics* (New Haven, CT: Yale University Press).

de Vaus, D. (2002), *Surveys in Social Research* (London: Routledge).

Fenno, R. (1978), *Home Style: House Members in Their Districts* (New York: Harper Collins).

Glaeser, E. L. and R. E. Saks (2006), 'Corruption in America', *Journal of Public Economics* 90(6): 1053–72.

Goel, R. and M. Nelson (2011), 'Measures of Corruption and Determinants of US Corruption', *Economics of Governance* 12(2): 155–76.

Heath, A., L. Richards, and N. D. de Graaf (2016), 'The sociology of corruption', *Annual Review of Sociology* 42.

Hitchcock, T. and R. Shoemaker (2006), 'Digitising History from Below: The Old Bailey Proceedings, 1674–1834', *History Compass* 4(2): 193–202.

Johnson, J. and H. T. Reynolds (2008), *Political Science Research Methods* (Washington, DC: CQ Press).

Keele, L. (2015), 'The Discipline of Identification', *PS: Political Science and Politics* 48(1): 102–6.

King, G. (2013), 'Big Data is not about the Data!' Talk at Harvard University Information Technology Town Hall, 28th October.

King, G., R. Keohane, and S. Verba (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press).

Laney, D. (2001), '3D management: Controlling data volume, velocity, and variety', *Application Delivery Strategies* (META Group, Inc.), 6 Feb.

Laver, M., K. Benoit, and J. Garry (2003), Extracting Policy Positions from Political Texts using Words as Data', *American Political Science Review* 97: 311–31.

Lazer, D., R. Kennedy, G. King, and A. Vespignani (2014), 'The Parable of Google Flu: Traps in Big Data Analysis', *Science* 343(6176): 1203–5.

Manski, C. F. (2013), *Public Policy in an Uncertain World: Analysis andDecisions* (Cambridge, MA: Harvard University Press).

Monroes, B. (2012), 'The Five Vs of Big Data Political Science. Introduction to the Virtual Issue on Big Data in Political Science', *Political Analysis* 19: 66–86.

Moser, Sir C. and G. Kalton (1971), *Survey Methods in Social Investigation* (London: Heinemann Educational Books).

O'Loughlin, B. and N. Anstead (2011), 'The Emerging Viewertariat and BBC Question Time: Television Debate and Real Time Commenting Online', *International Journal of Press/Politics* 16(4): 440–62.

Poole, K. and H. Rosenthal (2000), *Congress: A Political Economic History of Roll Call Voting* (Oxford: Oxford University Press).

Vanhanen, T. (1997), *Prospects of Democracy: A Study of 172 Countries* (London: Routledge).

Vanhanen, T. (2000), 'A New Dataset for Measuring Democracy, 1810–1998', *Journal of Peace Research* 37(2): 251–65.

Volkens, A., P. Lehmann, T. Matthieß, N. Merz, S. Regel, and A. Werner (2015), The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2015a (Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)).