


Watershed Resident Research

In this section you will learn how to:

- Collect and analyze quantitative data
- Sample watershed residents for surveys
- Design a watershed survey
- Conduct follow-up surveys
- Manage survey responses

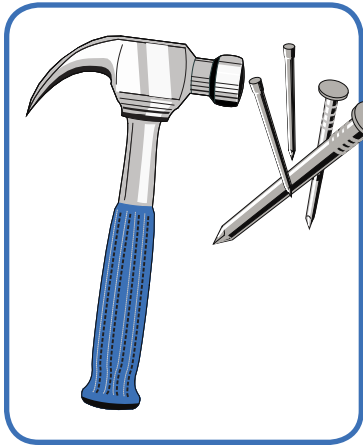


The first part of this guide includes sections on collecting and analyzing both quantitative data (survey work) and qualitative data (social mapping, interviews and listening sessions). This is for watershed action teams and coordinators who wish to gather useful information about the social landscape of their watershed.

By calling it a social landscape, we recognize that the people who live in an area have transformed the local ecosystem and watershed. While some features of the social landscape may be readily apparent by longtime residents or even casual observers, collecting systematic data about a watershed can provide objective support for the informal observations that are made about the communities and the environment.

Likewise, conducting research can reveal patterns that are unknown to us, because our familiarity with the area blinds us to seeing our communities in different ways. Social scientific research augments our understandings of our community.

Specifically, the community assessment tools described in this guide can be used by watershed action teams and coordinators to enhance community understanding of watersheds. Effective community assessments will allow watershed coordinators, SWCD commissioners, and others involved in the planning process to develop goals, outreach and education regarding water quality challenges based on the values and needs of the people living in the watershed.



Finding the Right Research Tool

Tools are designed for specific purposes; no one tool can do every job equally well. Although you can pound a nail with the broadside of a wrench, you'll get better results if you use a hammer. Likewise, using the right research tool in the right manner will prove to be more efficient and provide more accurate answers to the questions you would like answered.

Some research tools are designed for greater breadth of coverage, while others offer greater depth of knowledge. It is rare—and time consuming—for a study to cover both a large population and gain deep knowledge about the intricacies of individuals' behavior. Time and money constraints often dictate whether breadth or depth is possible, although there are ways to combine multiple tools in order to expand your understanding. In general, the kind of tool you choose to use can only give you one kind of depiction of the social landscape.

Broadly speaking, there are two categories of research tools: quantitative and qualitative methods. Quantitative research methods are tools designed to provide a measurement of behavior, attitudes or observable facts, usually expressed in some form

of statistics. Quantitative tools are an excellent way to reach a larger portion of a general population by asking specific questions that can be translated into a numerical expression. The hope is that, if collected in a systematic fashion, quantitative data can offer broad generalizations about a population. Two common quantitative methods are conducting surveys and censuses.

Qualitative research methods are tools designed to explore a limited number of cases more in depth in the hopes of offering reasoning and explanation of observable behavior, usually expressed in descriptive passages. Qualitative tools are an excellent way to understand motivations, rationale, attitudes and beliefs. Qualitative research methods are crucial in finding links between behavior and meaning. By examining a limited number of case studies in greater depth, qualitative data can give insight into why people do what they do, their self-explanations about their behavior, and how the beliefs they hold may impact what they do.

A common qualitative research method is conducting an in-depth interview. Qualitative research is suggestive of larger trends in a community, but cannot make systematic statements concerning how similar the limited number of cases investigated with qualitative research methods is to the rest of the population.

Quantitative and qualitative research methods complement one another. Effective survey questions, for example, can be written after conducting a number of qualitative interviews to determine areas of interest to examine in a larger population. Likewise, conducting interviews with a select sample of a population after survey collection can help interpret and provide greater context and background about participants' survey responses.

Quantitative research methods are tools designed to provide a measurement of behavior, attitudes or observable facts, usually expressed in some form of statistics.

Qualitative research methods are tools designed to explore a limited number of cases more in depth in the hopes of offering reasoning and explanation of observable behavior, usually expressed in descriptive passages.

This manual specifically covers three types of research tools: conducting a survey, social mapping and in-depth interviews. By carefully and effectively utilizing these three methods, you will be able to get a representative picture of the social landscape that is a watershed.

Random sampling selects participants in a study by chance.

Stratified sampling is used when you want to compare a group with some shared characteristic with other groups within a larger population.

Sampling and Research Design

In order for a study to be conducted, you must make a number of choices, all of which greatly impact the outcomes of your study. These choices include: what do you want to know more about, how will you go about collecting data, and what will be the ways you will evaluate the data you collect.

Knowing the Parameters of Study—Sampling a Universe

The first choice, what do you want to learn more about, sets the parameters of your study by limiting the population you will be investigating. This is known as the “universe” of your study. The parameters of what you wish to know about can be quite large. In fact, a universe can encompass thousands of people—think, for example, about national political polling, in which the universe is the entire number of likely voters in the United States. The larger the universe, the less likely that a study is going to be able to contact all members of that universe for questioning. In a nationwide poll, researchers would never be able to ask every eligible voter his or her political preference! Instead, pollsters make estimates by taking a representative sample of voters.

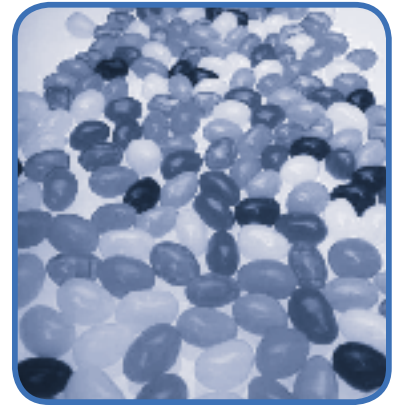
Imagine you have a large jar of jellybeans and you want to know how many of each flavor the jar contains. You could open the jar, dump the jellybeans all over the kitchen table and separate the flavors

into piles to count them. There would be so many cherry flavored jellybeans, so many licorice flavored jellybeans, and so on. However, you could take a sample handful from the jar and count the cherry flavored versus licorice-flavored jellybeans that are pulled from the jar.

You could then assume that the proportion of cherry and licorice jellybeans in the handful is more or less consistent with the proportion found in the jar as a whole. If the sample is large enough and collected in the correct way, the estimated proportion will be an accurate reflection of the proportion in the jar. However, if you choose too small of a sample there will not be enough variation in the handful of jellybeans to accurately describe the number in the jar. Likewise, if the licorice jellybeans are all settled at the bottom of the jar and you only take a handful from the top, your sample will be biased. Any conclusions you make will not reflect the universe you want to investigate.

Sampling—how you select the participants in your study—is a very important component of conducting research. You want your sample to be large enough and widely distributed to be accurate. In all cases, you need to know the total number of possible participants in your research. In the jellybean example, you would need to know how many were in the jar in the beginning. You can only select an accurate sample when you know the total number of possible choices from which you are choosing. If you are interested in assessing citizens' knowledge about water quality issues in a particular watershed, you will first need to know the total number of watershed inhabitants to choose a sample from them.

There are many kinds of sampling techniques that can be used to select a study's participants. There



are pros and cons to using any sampling technique, including monetary cost, time, reaching underrepresented groups and the type of research tool you want to use (for these purposes, mostly choosing between surveys and in-depth interviews). In this toolkit, two of the most widely used sampling techniques will be discussed: random sampling and stratified sampling. Both techniques are adequate for the types of research the majority of watershed action teams will conduct.

Random Sampling

As the name suggests, random sampling selects participants in a study by chance. It is also the easiest sampling method, if you have the right information. All the elements in the study universe are given a number. You choose how many samples you want and generate the set of random numbers that represent the subjects you will be investigating. For example, if you knew you had 450 households in your watershed and you wanted to do a phone interview with 20 of the households, you would label all the households from 1 to 450. Then, you would go to a random number generator, such as that found at the following website:



<http://www.random.org/sequences/>.

Enter the correct numbers (in this example, 1 is the smallest and 450 is the largest) and the program will generate a random list of ordered numbers 1 to 450. Because your sample size is 20, choose the first 20 numbers on the list and correlate these numbers to the corresponding households.

So, if the first number on your generated random number list were 322, you would make a phone call to the household that had been labeled “322.” In the event that household number 322 didn’t want to participate in the phone interview or the phone

number was unreachable, you would move down the list to the second number listed and find the corresponding household. You would continue to move down the list until you had conducted 20 phone interviews.

Stratified Sampling

Stratified sampling is a technique used when you want to compare a group with some shared characteristic with other groups within a larger population. To get a stratified sample, you would divide the universe into several groups based on their known proportion within the population. Stratified sampling is a great way to make sure your sample is representative of the larger population. It is also more economical when the research universe is large. Market research, for instance, often uses stratified sampling to test products across several categories such as age, gender and economic class.

For example, if you wanted to learn what students at a local high school know about water quality and you wanted to make sure your sample was representative across freshman, sophomore, junior and senior classes, you would use a stratified sampling technique. This way you could accurately compare the knowledge between different years in school. You would also make sure that you had a representative number of students from each class.



If the total number of high school students was 2200 and the number of students by class was: 304 seniors, 530 juniors, 634 sophomores and 726 freshmen, you would divide the number of students in a particular year by the total number of students in the high school, coming up with a percentage. So, 726 freshmen represent 33 percent of the total num-

ber of students in the high school (726 divided by 2200 is .33). Make sure you round up to the nearest percentage point. If you were going to sample 400 students, 33 percent of your sample, or 132 (400 multiplied by .33) should be taken from the freshman class. You would then follow the instructions for generating a random sample within the universe of freshmen students. This process would be repeated for all four classes in the high school.

Stratified sampling would be an ideal method to use when trying to gather information from disparate groups. For instance, if you are interested in self-reported water conservation behaviors and your watershed has an unequal distribution of rural and urban households, you would want to use a stratified sample. That way you would be guaranteed of not over or under representing either urban or rural households in your study.

Determining Sample Size

After you have chosen what sampling method to use, you will need to determine the sample size required in order to have accurate and statistically significant data. In general, the smaller the universe from which you are sampling, the larger the sample percentage is of that total universe. In order to calculate an accurate sample size, use the sample size calculator found at the following website:

<http://www.macorr.com/sample-size-calculator.htm>.

Keep the confidence level (how representative your sample will be of the total population) at 95 percent. The confidence interval is how much error is built into the sample, usually described in terms of percentage points deviating from the confidence level.

If your confidence level is 95 percent and your confidence interval is 3 percent, that means that the sample is accurate within a range between 92 and 98 percent. A lower confidence level and a higher confidence interval both adversely affect the representativeness of your sample. Finally, enter the population size and the calculator will generate the necessary sample size.

For instance, if the population is 10,000 households and you would like a 95 percent confidence level with an interval of 3 percent, your sample size would be 964. That means you would have to randomly select 964 households to take your survey in order to have a representative number of households.

A Word of Caution About Sampling

Regardless of the sampling technique used, there could still be error in your sampling because the universe from which you are drawing has built-in bias. Be aware from the beginning of your research plan how you define the parameters of your study can determine the outcome of your research.

For example, if you utilize the area phone book as your universe of residents in a town from which to draw your sample, be aware that this parameter does not reach all potential town residents. Missing from your sample would be all town residents who do not own a land line phone and those who, for the sake of privacy, prefer to have an unlisted telephone number. Likewise, if you wish to sample members of the area Ducks Unlimited chapter and are using a generated list of members, your sample will only be as accurate as the most updated roster.

The most accurate universe for the purposes of conducting research on the social landscape of a



Be aware from the beginning of your research plan how you define the parameters of your study can determine the outcome of your research.

watershed would be to assemble a mailing list of all residents and landowners within a given watershed. Our team worked with five watershed groups to develop and test the set of research tools included in this guide. We surveyed all of the residents living in four of the five watersheds. Unfortunately, we were unable to do this for Silver Creek due to funding constraints.

Silver Creek Watershed has approximately 6,900 residents if we include the whole city of Cresco. While only partially in the watershed, the entire city was included because of shared drainage issues. We created a mailing list for the watershed by compiling landowner lists from Howard and Winnebago counties and a mailing list from the city of Cresco. Due to the large size of the Cresco mailing list and limited funding to conduct the assessment, we decided that every fifth resident should be entered into the survey mailing list. The list was still very large, so the final list was determined using every other entry and equates to approximately 10 percent of the watershed population plus the city of Cresco, or 690 residents.

So funding for the assessment can—and often does—dictate whether you sample everyone in the watershed or need to generate a random sample. It is important to treat the mailing list as a living document. This means that the list will need to be updated regularly as addresses change, people move or pass away. Detailed instructions on how to gather the addresses for a watershed, generate a random sample from the list using a simple computer program and generating a mail merge can be found in Appendix 2: Assembling a Watershed Mailing List.

Designing a Survey

Surveys are an excellent method to gather specific information that can later be quantified and entered into a database so that the information can be expressed in statistical form. Surveys are typically questionnaires that gather demographic data and ask specific questions pertaining to the area you want to know more about.

All surveys should include a section that collects demographic information. Demographic questions should reflect the characteristics of the sample population about which you are interested. Common demographic variables include: age, income, residence (rural vs. urban, township, county, etc.) marital or relationship status, number of dependents or number and age of people living in the household, education level, gender, and ethnicity (not “race”). Be sure to include a space that allows respondents to write in how they self-identify for ethnicity.

Survey questions can be factual (e.g. On average, how many times a month do you fertilize your lawn?), attitudinal (e.g. In your opinion, who should be the most responsible for protecting water quality in your community?) or hypothetical (e.g. How likely are you to get involved in working with your local watershed protection association?).

When designing a survey questionnaire, it is important to remember that whatever type of question is asked (factual, attitudinal, or hypothetical), the answer must be translatable into a numerical expression. Sometimes this is built into the question itself. For example, in asking a factual question like, “On average, how many times a month do you fertilize your lawn?” the reply will already be a number. Questions that ask frequency (how many times do you...), quantity (how much do you...) and size

It is important to remember that whatever type of question is asked the answer must be translatable into a numerical expression.

(how big/small is...) will all be answered in this way. Common questions of this kind include:

How long have you lived in your area?

What is your age?

What is your annual income?

How many acres do you currently farm?

In all these cases, the questions will be answered with a number expressing the number of years, number of dollars earned, and number of acres farmed.

Unlike factual questions, attitudinal and hypothetical questions are not expressed in numerical value and thus need to be “translated” into a number.

The most common way of doing this is to limit the range of responses, or what is called a forced choice format, where you “force” the participant to make a choice. These can be yes/no questions, ranking questions and scaled questions.

Yes/No Questions

Yes/No questions are those that require the survey participant to reply with a simple yes or no answer. These questions should be clear and unambiguous as to the type of response. In other words, there is no middle ground in the reply, it either is or it isn't. A quality yes/no question would be:

Do you feel that your home drinking water is safe to drink? YES NO

This is an effective yes/no question because, although it is asking the opinion of the respondent, there are clearly only two possible answers. Participants would circle either “yes” or “no.”

An ambiguous, ineffectual yes/no question would be:

Are you a homeowner or a renter? YES NO

This question produces an ambiguous response. We do not know if the respondent is either a homeowner or a renter, only that they are one or the other (and, frankly, if they do not own their own home or rent where they live, what other option would there be?). We could transform the above question into a much more effective one by altering it in the following manner:

Are you a homeowner? YES NO

OR

Do you currently pay rent where you live?

YES NO

Either of these questions will garner clearer responses.

A variation on the yes/no question is to provide a set number of possible responses, of which the respondent must choose. Such questions can express a range of attitudes or opinions from which the respondent can only choose one. In the sample question below, participants are asked to circle one answer from a list that expresses a spectrum of soil erosion from "none" to "a lot." Notice that "don't know" is included as a possible response.

Do you have any soil erosion on your property?

NONE A LITTLE MODERATE

A LOT DON'T KNOW

Likewise, such questions can include a number of possible answers from a list, in which the participant can choose multiple responses. The responses are not

ranked in any way and hence are equally weighted. In the sample question below, respondents are asked to choose up to three possible answers.

In your opinion, which of the following are most responsible for the existing pollution problems in rivers and lakes *in Iowa*? (CHECK UP TO 3 ANSWERS)

- Agriculture crop production
- Erosion from roads and/or construction sites
- Wastes from urban areas
- Industry
- Wild animals/pets
- Livestock and/or poultry operations

Again, in this example, the participants are choosing their answers and, because they are not asked to rank them, they will all be treated as of equal value when it comes time to statistically evaluate the responses. In other words, because of the format of the question, you will only know which of the responses get included most frequently as being responsible for water pollution. However, you will be unable to ascertain which response is viewed as being the most responsible. In order to know this, you would have to ask respondents to rank their choices.

Ranking Questions

Ranking questions, as the name implies, asks respondents to assign relative value to a list of possible answers. In such questions, the survey participant is sorting answers from highest to lowest. Typically, this is expressed with 1 being the least

likely or least frequent response and higher numbers represent more likely or more frequent responses. By making a few minor adjustments to the wording of the sample question above, it is transformed into a ranking question:

Please rank the following as being responsible (1 being the least responsible and 6 being the most responsible) for the existing pollution problems in rivers and lakes in Iowa.

Agriculture crop production _____

Erosion from roads and/or construction sites _____

Wastes from urban areas _____

Industry _____

Wild animals/pets _____

Livestock and/or poultry operations _____

By including a ranking question such as this, you will be able to determine the relative value of participants' perceptions of water pollution. You would be able to compare the perceptions of different demographic categories to see if differences existed. For instance, if on the same survey you asked whether or not the participant lived in a rural or urban setting, you could then compare if rural residents' perceptions of water pollution differed from urban residents' perceptions.



A Word of Caution About Writing Effective Survey Questions

All questions should be tested before going through the time and expense of administering a survey. What may seem perfectly clear to the writer of a survey question may not be interpreted in the same way by a participant. It is recommended that you pre-test the survey with seven individuals who are not directly involved in the project, are varying ages, sex and rural/urban residencies.

In addition, all survey questions should have clear instructions embedded in the question so that participants know how to answer the question. These instructions include comments like “Check all that apply,” “Choose 1 answer,” “Choose 3,” and “Rank answers from 1 to 5, 1 being least likely and 5 being most likely.” Another reason to pre-test questions is to make sure the instructions are clear and easy to follow.

Included in this research toolkit is the standard questionnaire for investigating water issues in Iowa that the Heartland Regional Water Coordination Initiative developed, used statewide and throughout USEPA Region 7 for a number of years (Appendix 1). The questions on this survey have been tested for their clarity. We recommend that watershed groups incorporate these questions into their own study because they have been vetted and because they will allow watershed groups to compare their findings against other watersheds in the state. However, we also encourage watershed groups to formulate their own questions for issues specific to their communities.

Well written, forced choice survey questions are easier and faster for respondents to answer, which often result in a higher response rate. If there is a

specific issue or area that you wish to know more about or that is a new area of concern for you, forced choice survey questions are not the best method. Instead, conducting an in-depth interview with open-ended questions will allow you to explore an area as well as ask for immediate clarification and needed background in order to understand the context of the response. Open-ended questions are covered in the section on qualitative interviewing on page C-1.

Survey Response Rates

A high percentage of people who respond to a survey, also known as a response rate, is crucial if the survey's results are to be taken seriously and said to be representative. When a survey's response rate is low, it lessens the validity of the findings as well as provides an inaccurate overall picture of the social landscape.

A survey response rate is different than the sample size you determined was needed in order for the survey to be representative of a larger population. You can obtain the response rate by dividing the number of participants submitting *completed* surveys (typically 80 percent or more of the survey questions answered) by the number of participants the survey was sent to. Thus, if 300 people were mailed a survey and 75 completed and returned the survey, the response rate would be 75 divided by 300, or 25 percent. A 25 percent response rate is a poor result; you should aim for the highest response rate possible.

Acceptable response rates vary depending upon how the survey is administered, as outlined in the chart below. *You can make stronger claims to the representativeness of your sample as the response rate gets closer to 100 percent.*

Table 1: Standard acceptable response rates

Type of Survey	Targeted Response Rate
Mailed Survey	40% Adequate 60% Good 70%+ Very Good
Phone Survey	80% Good
Email Survey	40% Average 50% Good 60%+ Very Good
Online Survey	30% Good
Face-to-Face Survey	80% Good

Table 2: Watershed-based Community Assessment acceptable rates for mailed surveys

Type of Survey	Targeted Response Rate
Mailed Survey	20% Poor
	30% Adequate
	40% Good
	60% + Very Good

The second chart is a reasonable response rate for watershed-based community assessments based on the amount of time and resources available to conduct the survey. It is important to get at least a 50 percent response from a mailed survey when the purpose of the survey is to measure effects or make generalizations to a larger population. However, a lower response rate is acceptable if the purpose is to gain insight and direction for outreach and education, as is the case in the watershed-based community assessment survey. Please note that if you get a response rate of less than 30 percent, this indicates a lack of engagement on the part of the residents in your watershed and could hinder the overall success of a watershed improvement project.

Low response rates are a continuing problem for groups. Our society has been inundated with questionnaires and surveys via the U.S. mail, email, or by phone. Some people simply refuse to participate in surveys, while others, for a wide range of reasons, cannot participate. Still, a well-designed survey, coupled with techniques to elicit response, can help guarantee a good response rate.

Sending Out a Survey and Survey Follow Up

In order to know which participants have returned surveys and who to send follow-up information to, it is necessary to track the surveys. Tracking is typically accomplished by assigning each potential participant a number and putting that number somewhere on the survey itself. Because some participants prefer not to be identified to all, it may be best not to make the assigned number obvious.

Investigators must remember that all individual participant information must be kept confidential. While it is standard to talk about aggregate findings (for example, 23 percent of respondents were males over the age of 65), **in no case should you reveal the information of a particular respondent. This is a matter with legal ramifications.** There must be no way for anyone outside of the research team to ever connect a participant to his or her survey responses.

The importance of follow up cannot be overstated. Ensuring a high response rate typically means attempting contact with non-responding participants at least twice after the initial contact attempt. It is also important to prime the participants by letting them know that a survey will be coming and informing the general public about the nature

It is important to prime the participants by letting them know that a survey will be coming and to inform the general public about the research.

of the research and the importance of receiving each participant's answers. This can be achieved through articles and press releases sent to local newspapers and area radio station interviews.

For example, the Sac and Carroll County Soil and Water Conservation Districts sent a letter to the Black Hawk Lake watershed residents in advance of mailing the survey; Badger Creek Lake watershed put an article about the survey in their newsletter prior to survey mailing. This was not done in the two other surveyed watersheds. Black Hawk Lake watershed had the largest response rate (45 percent to-date at time of printing) of the four watersheds surveyed and Badger Creek Lake had the second highest response rate (34 percent). The simple act of letting people know that a survey will be coming can significantly increase the number of responses.



After letting the residents know that a survey is coming, it is customary to send the survey along with a letter explaining why the survey is being conducted, how to complete the survey and how the anonymity of participants' responses are guaranteed. Within two to four weeks after the first mailing, a reminder should be sent to the entire mailing list. The reminder (a postcard or a letter) should thank those who have already sent back the survey and remind others to do so. It should include the investigator's contact information as well as information on how to obtain a replacement survey if the first survey was lost or didn't arrive. A second survey should be sent to those who haven't responded (or those requesting another copy) within six to eight weeks after the first mailing. Depending on your response rate after two mailings, a third mailing may be necessary to achieve appropriate response rates.

How to Manage Survey Responses

After conducting the survey and receiving responses, how do you make sense of the information you have collected? Generally, survey responses must be tabulated in order to calculate simple frequencies (i.e. how many participants responded in a particular way) and make comparisons across types of categories.

There are a number of statistics software programs on the market, such as Systat and SPSS. These are typically more advanced than what most watershed groups will need in order to analyze their survey material. Unless someone in your group already has familiarity with a statistics program, it is recommended that you manage the survey data in a simple spreadsheet program like Excel, which is usually included with Microsoft Office software.

Excel is quite useful for recording data from surveys. Many people already know how to use Excel, so special training is typically not needed. Enter information from each completed paper survey into an Excel spreadsheet as they arrive by mail. Number the printed survey's first page in the upper right hand corner with the spreadsheet row number into which the survey data was entered. This is done so that the paper survey may be consulted later to double check accuracy, if needed. For the purposes of the following sample instructions, the first page of the "Water Issues in Iowa" survey will be used; shown in Appendix 1.

For the purposes of entering survey data, the rows (labeled down the left side 1, 2, 3, etc.) will be the individual survey responses received from each participant. The columns (labeled across the top A, B, C, etc.) will represent the questions on the survey.



Thus Row 1, Column A will represent the answer to question number 1 by the first survey respondent that you are entering.

When setting up your survey data spreadsheet (figure 1), first enter the question categories as each column. The first row of every spreadsheet includes all of the question labels (and not numerical values). Using the “Water Issues in Iowa” survey, column A represents the participant survey ID number that

	A	B	C	D	E	F	G	H	I
1	SurveyID	Question1	Question2A	Question2B	Question2C	Question2D	Question2E	Question2F	Question2G
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									

Figure 1: Excel spreadsheet illustrates the file set-up for recording survey responses.

was generated as part of the sampling technique used. Column B, then, represents the category “Question #1.” You only need one column to represent the first question because of the format of that question. Question #1 forces respondents to only choose **one** answer; thus when entering the replies into the spreadsheet, you only have to record that one answer.

Not all questions are as straightforward as Question #1, however. Question #2, for instance, allows respondents to check **more than one** reply. In such instances, you need to be able to record how the respondent replied to each answer in Question #2. When setting up your spreadsheet, you need to take into account that Question #2, because it allows multiple replies, must be counted as multiple questions.

In this example, Question #2 allows up to seven different replies, so Question #2 is represented on our spreadsheet as seven different questions, labeled #2A through #2G; thus, Column C represents Question #2A, Column D represents Question #2B all the way until Column I, which represents Question #2G. Again, if a question allows for more than one possible choice, each choice must be treated as a separate column in the spreadsheet.

The same rule applies to other questions that ask respondents to rank or evaluate from a range of possible answers. Question #6 on the “Water Issues in Iowa” survey asks respondents to what degree they suspect whether eight separate conditions are affecting their water quality. Like in the example of Question #2, each of the eight conditions must be entered as separate columns in the spreadsheet.

After the spreadsheet has been set up with the columns representing each question category, you need to assign values to each answer. Surveys need to be “translated” into numerical values so that they can be easily entered into Excel. In order to do this, take a blank copy of the survey and assign numerical values to the forced choice replies.

For instance, refer to Question #1 of the “Water Issues in Iowa” survey:

1. What is the best definition of a watershed?

(CHECK ONE BOX)

- A structure that stores water
- An area of land that drains to a common body of water
- A basin to hold extra water to prevent flooding
- An underground water supply

Alongside each answer you would mark a number to represent that answer choice (indicated in blue):

1. What is the best definition of a watershed?
(CHECK ONE BOX)

- 1 A structure that stores water
- 2 An area of land that drains to a common body of water
- 3 A basin to hold extra water to prevent flooding
- 4 An underground water supply

On the Excel spreadsheet, if the first survey you are entering into the program replied with the second choice (A watershed is an area of land that drains to a common body of water) you would type a number "2" into the corresponding row and column: Row #2, Column B. It is Row #2 because it is the first survey you are entering (as Row #1 contains your headings) and Column B because that column corresponds to "Question #1."

	A	B	C	D	E	F	G	H	I
1	SurveyID	Question1	Question2A	Question2B	Question2C	Question2D	Question2E	Question2F	Question2G
2	1	2							
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									

Figure 2: Excel spreadsheet showing the first survey response for question 1.

If a question allows for more than one choice to be made, you need to code each possible choice. This is why when creating the spreadsheet (as indicated above in figure 2) you assigned multiple columns to questions that allowed more than one reply to be chosen from a list of possible answers. Take, for instance, Question #2 from the "Water Issues in Iowa" survey:

2. Where do you get your drinking water? (CHECK ALL THAT APPLY)

- Private well (individual well or well that serves fewer than 15 residences)
- Rural Water system
- City water system
- Purchase bottled water
- Don't know

Because you need to know if a respondent replied to more than one choice, you treat each choice as a yes/no question. Assign a standard value of "1" for YES and a standard value of "2" for NO, consistent throughout the spreadsheet. If a respondent replied that they get their drinking water from a private well, you would indicate this in the spreadsheet as a YES reply (the value of "1").

Under Column C (representing Question #2A on the spreadsheet) you would type a "1" to indicate a positive response. If a respondent left the box blank for the reply "Rural Water system," you would indicate this as a negative reply on the spreadsheet. Under Column D (representing Question #2B on the spreadsheet) you would type a "2" to indicate that

	A	B	C	D	E	F	G	H	I		
1	SurveyID	Question1	Question2A	Question2B	Question2C	Question2D	Question2E	Question2F	Question2G	Question2H	Question2I
2		1	2	1	2						
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											

Figure 3: Excel spreadsheet illustrating how to enter in the first survey respondent's answers to question 2.

the respondent does not get drinking water from an individual system (figure 3).

Finally, questions that ask respondents to rank or evaluate a range of possible replies can be treated as a kind of forced choice response within each criteria being evaluated. This is why you entered each of the eight conditions as a separate column in Question #6 of the “Water Issues in Iowa” survey. Thus, use a combination of approaches as outlined for Question #1 and Question #2 above.

Count each condition (high bacteria counts, fertilizer/nitrates, etc.) as a separate column in the spreadsheet. Across the top, assign each assessment of knowledge (Know, Suspect, Not A Problem, Don't Know) a number that would represent its value in the spreadsheet, as indicated below by the blue numbers across the top:

6. Do you know of or suspect any of the following conditions are affecting water quality in your area?

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
	<i>Know</i>	<i>Suspect</i>	<i>Not a Problem</i>	<i>Don't Know</i>
High bacteria counts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fertilizer/Nitrates	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Heavy Metals (e.g., lead, arsenic)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hardness (e.g., calcium, other minerals)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pesticides	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Animal waste	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Septic systems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pharmaceuticals (i.e. antibiotics, personal care products)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

You would then take each condition and enter the corresponding number into the spreadsheet, as illustrated in the next example (figure 4).

Example survey responses:

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
	<i>Know</i>	<i>Suspect</i>	<i>Not a Problem</i>	<i>Don't Know</i>
High bacteria counts	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fertilizer/Nitrates	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Heavy Metals (e.g., lead, arsenic)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Hardness (e.g., calcium, other minerals)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Pesticides	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Animal waste	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Septic systems	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Pharmaceuticals (i.e. antibiotics, personal care products)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

M	N	O	P	Q	R	S	T
QuestionbA	QuestionbB	QuestionbC	QuestionbD	QuestionbE	QuestionbF	QuestionbG	QuestionbH
1	2	4	4	2	1	3	4

Figure 4: Excel spreadsheet indicates the first survey respondent's answers for all the choices in Question #6.

Questions that ask for a numerical response need not be coded. The numerical value can be directly entered into the spreadsheet. Thus, for a question that asks, "How long have you lived in your area?"

you would enter the number indicated on the survey by the participant.

Finally, in the event that the survey allows for an open-ended response, these responses can still be coded and assigned a numerical value. For instance, if Question #2 allowed respondents to check “other” and included a space for respondents to write in another answer, you would assign each write-in a numerical value. You could choose to aggregate some responses that were similar to one another as part of the same numerical value.

Thus, if one respondent checked “other” and wrote in that they get their drinking water from a swamp and another respondent checked “other” and wrote in that they got their drinking water from a marsh, you could code both swamp and marsh with the same numerical value because of the similarity of the two responses.

Every question in a survey must be coded in order for the information to be entered into the spreadsheet. If the survey is coded incorrectly or not coded completely, it will skew the results when you tabulate frequencies and make graphs from the information gathered.

Likewise, it is paramount that the information from the surveys be entered correctly and carefully into the spreadsheet to eliminate as much error as possible. As you enter data from the surveys, you need to be alert and aware of possible data entry errors.

To aid in interpretation of data with Excel software, please see Appendix 3: Data Interpretation with Excel Software.