

Data Interpretation with Excel Software

Various Excel procedures may be used to help you interpret and make sense of your survey data. These procedures are referred to as descriptive statistics and graphs (e.g., pie charts, bar charts, histograms and frequencies). These very important data interpretations provide investigators with the information needed to make informed decisions about the issues and concerns their research was designed to address.

While there are certainly many different statistical procedures that can be utilized, simple descriptive statistics and graphs often provide enough information to address many issues. In fact, frequencies and graphs are often the preferred approach to sharing data with community stakeholders.

Tabulating Frequencies

In order to put into graphic form the data you have collected from your survey, you will need to first calculate the number of different answers for each question, known as the frequency.

Frequency is the number of times a particular response to a particular question occurred throughout all the surveys collected. For example, you would want to know how many respondents answered “yes” and how many answered “no” to Question #3, “Do you feel that your home drinking water is safe to drink?”.

You could go through the spreadsheet and do a manual count of 1’s and 2’s (representing your coding of “1” for “yes” and “2” for “no”), but this could take a long time and leaves plenty of room for human error, especially if the number of surveys received reached the hundreds. Instead, Excel can do the counting for you, provided you enter the right instructions, or function, to do so.

In order to add up all the different entries in each question, you need to first go to the row below the last survey entered. In the small sample, there are 11 total rows with data entered into them; leave a space (Row 12) and begin with Row 13 as the place you would insert counting instructions for

Excel. Click on the corresponding column you wish to count.

For the first example, you will discover the frequency of replies for Question #1, represented by Column B in your spreadsheet. Click on the space Row 13, Column B. Once clicked, the box will be outlined in blue. In order to insert the “instructions” go to the formula builder, which may be located in different locations depending on your computer operating system. You may see a “Formulas” tab across the top row of your Excel program (between Page Layout and Data). Click on this tab, and the *fx* Insert Function button is located at the top left. Instead you may have a floating toolbox on the right hand side of the screen (it will say “formatting palette”); if this is the case, click on the picture of the “*fx*.” The “formatting palette” now changes its name to “formula builder.”

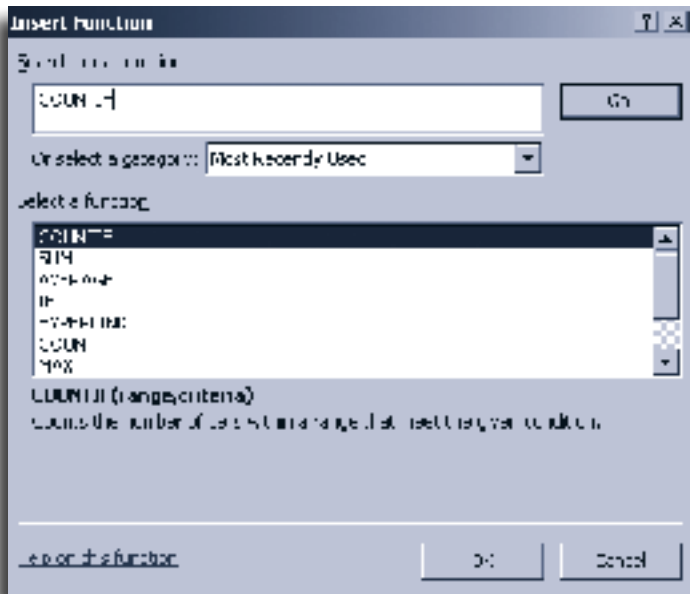


Figure 3-1. Excel Insert Function window that will tabulate columns automatically.

“range” and “criteria.” You will be entering information into the range and criteria fields as in figure 3-2.

Range represents all the information you want included in the count. Click on the blank field marked range. Because you are interested in Question #1, your range is all the entries under question #1.

You will be using a special formula to do the counting for you, called COUNTIF. Type COUNTIF into the search query at the top of the formula builder (see figure 3-1).

When COUNTIF is displayed in the box below the search query, double click on it. A description of the COUNTIF formula will appear below the display box and under the description a space marked “Argument” with two blank fields for

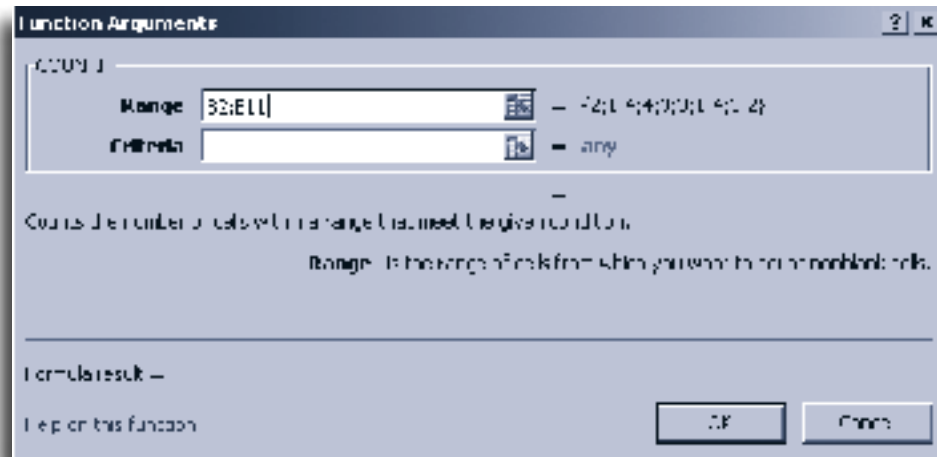


Figure 3-2. Excel Function Arguments window describing the COUNTIF function.

With your cursor click on the first entry of data under Column B, which is Row 2. Hold the click and drag the cursor down to include all the replies in Column B—this goes to Row 11 in the sample spreadsheet. Notice that in the field marked “range” is now entered the following sequence: B2:B11. This means the “range” you are interested in counting includes all information entered from Column B, Row 2 to Column B, Row 11. Instead of using the mouse to click on the spreadsheet, you could also have typed in the exact range expressed as Column #: Row #, separated by a colon.

Next, click on the criteria field. This represents the value you want counted. If you wanted to know how many people responded to Question #1 with the first answer, you would enter “1” in the criteria field. The number “1” represents the code you have assigned to the first possible reply to Question #1. It is crucial that the number entered into the criteria field is surrounded by quotation marks. This tells the formula to only look for that exact number.

Press the “enter” key and a new number appears in the space—in the sample spreadsheet (figure 3-3), the number 2 appears, representing that there were two respondents to the survey who replied to Question #1 with the first answer.

You must do this count for all possible answers to Question #1. Because there were four possible answers, you must count for all four—represented in the coding as the numerical values 1 through 4. In other words, you

	A	B	C	D	E	F	G	H	I
1	SurveyID	Question1	Question2A	Question2B	Question2C	Question2D	Question2E	Question2F	Question2G
2	1	2	1	2	2	2	2	2	2
3	2	1	2	2	2	1	2	2	2
4	3	4	1	2	2	2	2	2	2
5	4	4	1	2	2	2	2	2	2
6	5	1	2	1	2	2	2	2	2
7	6	1	2	1	2	2	2	2	2
8	7	1	1	1	2	2	2	2	2
9	8	4	2	2	2	1	2	2	2
10	9	1	2	2	2	1	1	2	2
11	10	2	2	2	2	2	2	2	1
12									
13									
14									
15									
16									
17									

Figure 3-3: Excel spreadsheet example totalling responses for question 1.

must now calculate the frequency of survey participants responding with the second, third, and fourth possible replies.

In order to do this, first place your cursor directly below the place you just entered the formula to count the frequency of the first reply (in the sample, this is Column B, Row 14). Follow the same instructions as above except this time in the criteria field enter "2" instead of "1" to count the frequency of the second response.

When finished, place the cursor directly below this number (Column B, Row 15) and repeat with the criteria field entered with a "3." Place the cursor in the next cell (Column B, Row 16) and repeat for a final time with the criteria field entered with a "4." In the sample spreadsheet, the following are the frequencies for Question #1:

1st Reply: 2 3rd Reply: 3
 2nd Reply: 2 4th Reply: 3

For each question that you have coded with a numerical value, you will have to determine the frequency of responses in order to depict the data in a graphic or chart.

Once you have determined the frequency of responses, you can also express any one response as a percentage of all responses to a question by dividing the number of a particular reply by all the replies in a question.

For example, in the sample spreadsheet for Question #1, the number of survey participants that responded to the question with the first answer was 2. The total number of surveys collected was 10. Divide 2 by 10 and you arrive at .2, or 20 percent. In other words, 20 percent of survey respondents replied to Question #1 with the first answer. You are now able to graph the frequency of responses to the first question.

Using Graphs to Present Data

Since studies show that our minds retain only 10 percent of what we hear, but 50 percent of what we see, graphs are a powerful tool for presenting information. Graphs provide “pictures” of the data and allow investigators to identify various responses from their survey participants. Since there are many types of graphs, we must choose the best possible graph for the message. This section will provide a few examples of graphs with both strengths and weaknesses.

For watershed-based community assessments, data are reported in written form, usually by numbers either in tables or entered into the original survey to indicate the responses. A properly prepared graph can help report the data in a more visual form. Whether the need is to inform or to persuade, graphs are an efficient way to communicate. Graphs can be a great help in the presentation of information and also in the analysis of data. However, graphs can be time-consuming, so it is important that you are selective in which data you represent graphically and what data is presented in tables or text.

Above all else, graphs need to be understandable to the intended audience. Do not make your graphs unduly complex. A graph must attract and hold the attention or the message of the graph will not be well received by the audience. A good graph establishes or expands interest in the data it represents. A graph can make the message more explicit and help the audience think clearly about the issue at hand. Following are examples of a few good graphs and explanations as to why they effectively represent the data.

Finally, make certain you supply an interpretation for the graph whether you are placing it in a report, a handout or showing it in a PowerPoint presentation. Every graph should have an explanation. If you don't know what the graph says, then don't use it.

Comparison #1

Pie graphs are a useful tool for visually representing parts of a whole. The same data are presented on both graphs here. Note that both graphs are appropriately labeled with a title, and the pieces of each pie are given a descriptive label to aid the viewer's understanding of the data. However, numerical percentages are not included on the top pie graph (Figure 3-4), leaving the viewer to guess what the respective sizes of each piece represent. Humans are notoriously bad at visually comparing areas (sizes) in a graph like this, and this problem is further amplified by the use of 3-D effects. In the graph in figure 3-4, note how the 3-D effects add major visual distortion. While the two of the pieces are actually identical in size, the piece in the foreground appears substantially larger and can subsequently

skew how the data is interpreted. For the most effective pie graphs, include a title, stick to simple 2-D graphs and be sure to label each category (piece of the pie) with its descriptive label and corresponding numerical value as illustrated in figure 3-5.

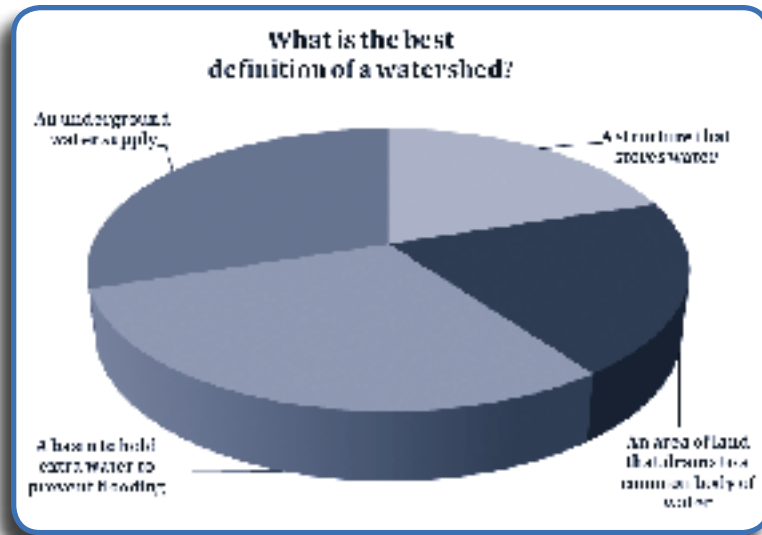


Figure 3-4: Example of a poor pie graph, above.

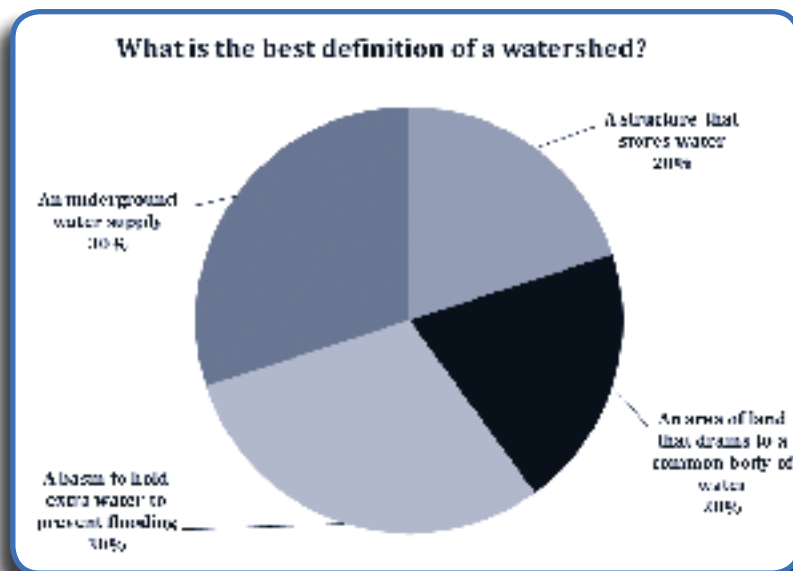


Figure 3-5: Example of a good pie graph, at right.

Comparison #2

Both of the bar graphs below (Figures 3-6 and 3-7) have several strong features, including a title, label on the x-axis (Response Frequency), and descriptive labels for each of the responses on the y-axis. However, note the differences in the values on the x-axis between the two graphs. The graph on the left goes up to a Response Frequency of 5, wasting a significant amount of space, while the graph on the right has been adjusted to a maximum value of 3.

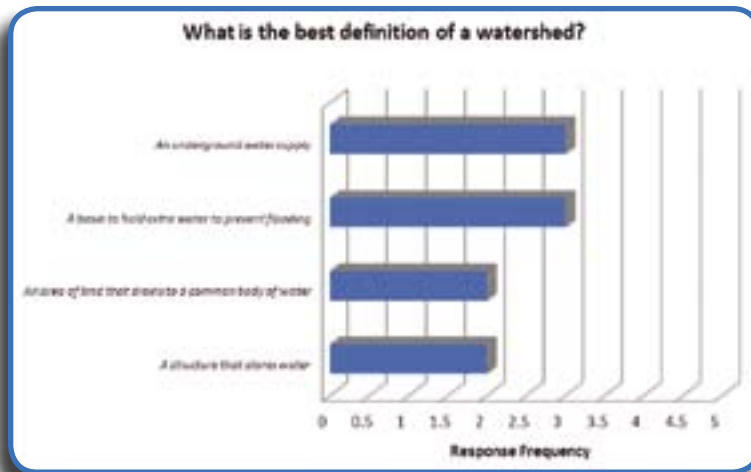


Figure 3-6: Example of a poor bar graph

you could not have 2.5 responses. There are too many decimal places for this data set; this can also be adjusted in the Axis Options menu.

Finally, resist the temptation to include fancy 3-D effects! Note the visual distortion in the left-hand graph; the top two responses have a response frequency of 3, yet the 3-D bar appears to be at a value closer to 2.8, making the graph difficult to read and interpret. For the most effective bar graphs, include a title, descriptive labels on the x-axis and y-axis,

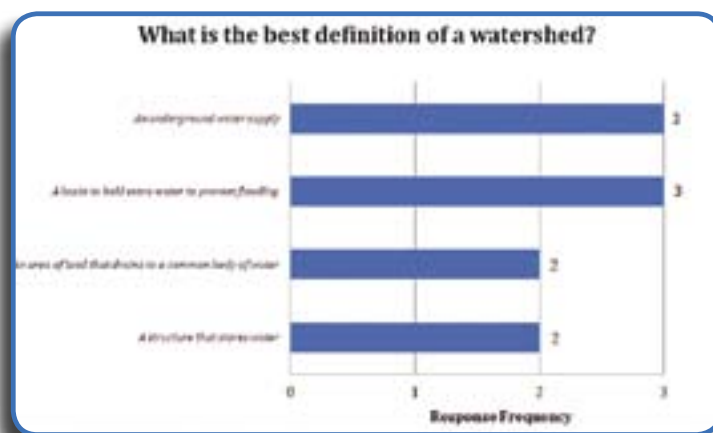


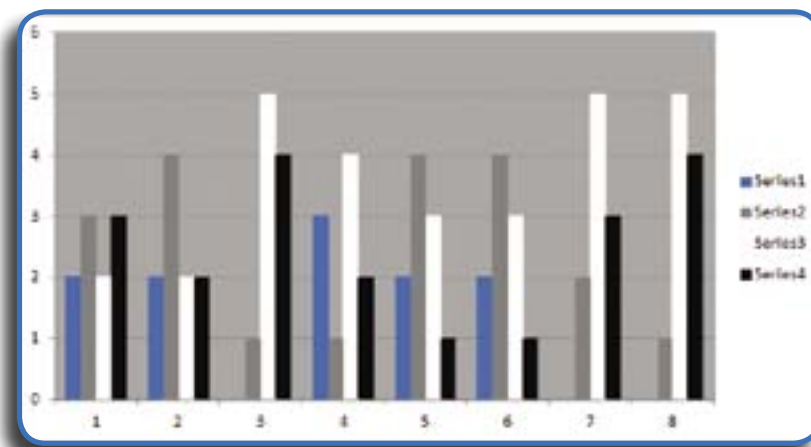
Figure 3-7: Example of a good bar graph

use the Axis Options menu to adjust the axes appropriately for your data

set, and stick to simple 2-D graphs for the greatest clarity and effectiveness as illustrated in the bar graph in figure 3-7.

Comparison #3

Upon selecting data and creating a column graph in Microsoft Excel, the bar graph in figure 3-8 is representative of what your output may look like. As it currently stands, the x-axis, y-axis and legend are each labeled with simple default numerical labels. Without descriptive axis and legend labels, the graph is largely unclear and meaningless for the user. Furthermore, some versions of Microsoft Excel provide a default gray background to all graphs created. This poor color contrast further reduces any visual effectiveness



this graph may have had. This is a prime example of a graph that includes too much information without proper labeling, leaving it rather useless overall.

Figure 3-8: Excel's default bar graph with initial data entered.

A number of improvements were made to the graph in figure 3-9, including the addition of a title, appropriate axis and legend descriptive labels, and the plot area has been changed to white. This is now a much more effective graph! However, the graph is quite busy. With a large amount of data such as that collected in the "Water Issues in Iowa" survey, it is tempting to cram too much information onto one graph. More isn't always better, as the audience can easily get lost in the data and lose sight of any take-home message. It's important to think about what your goals and messages are in presenting data graphically.

Perhaps your local water body is impaired for bacteria. So one of the more interesting pieces of information from the above data set is that related specifically to high bacteria counts. Creating a simple graph that illustrates

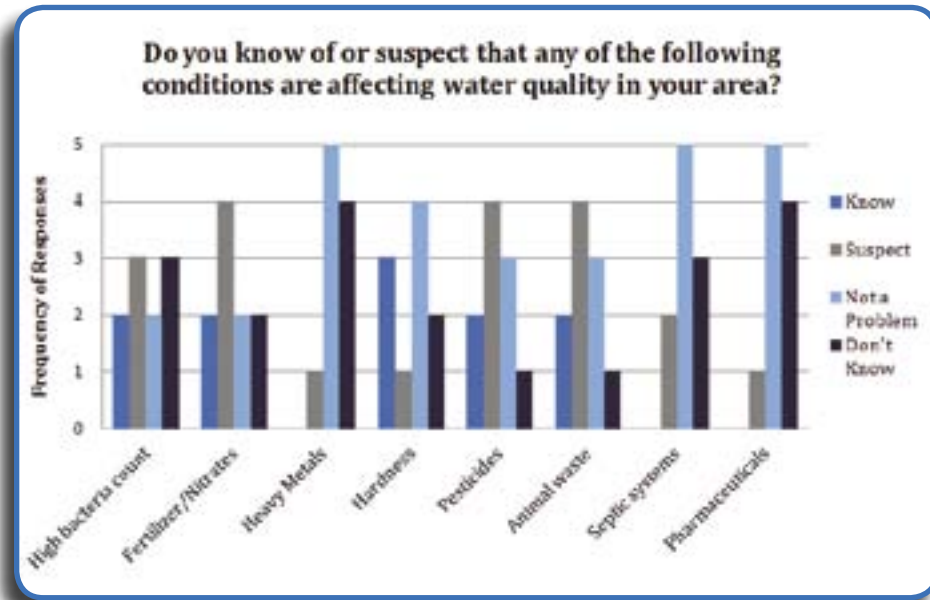


Figure 3-9: Excel bar graph with legend, title and background changed.

residents' perceptions of high bacteria counts as a source of water quality challenges is pertinent and useful. Rather than presenting data related to eight different possible water quality conditions, focusing on 1-3 that are most pertinent to your watershed, its impairments, land uses, and challenges would be much more effective, as shown in figure 3-10. Choosing which data to represent graphically can be just as important as creating an effective graph with a title and proper descriptive labels.

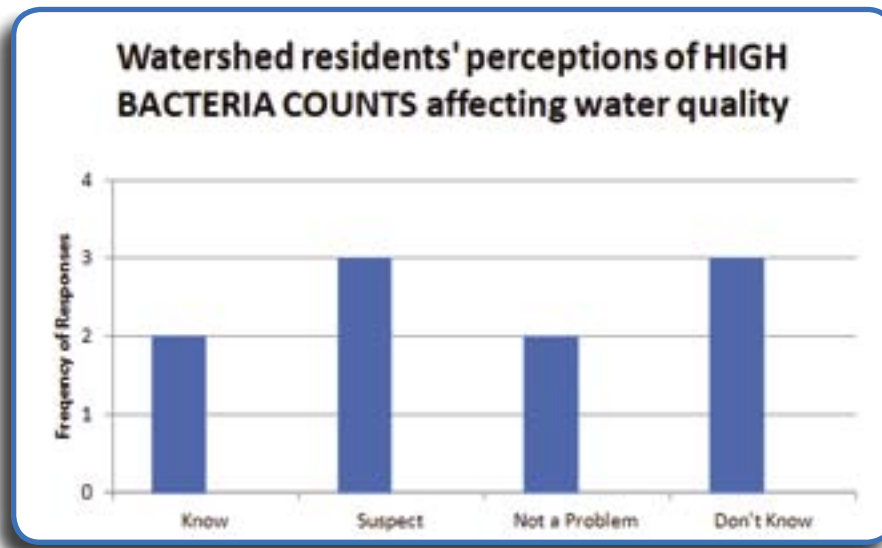


Figure 3-10: Simplified bar graph in Excel illustrates data clearly.

Graphing Your Data

Graphing data is an important procedure for presenting your data to the public. Graphs can be generated after all the data has been entered into the Excel spreadsheet and the frequencies of responses have been calculated. Not all graphs are equally good at representing all varieties of data. Some, like pie graphs, are good representations of frequencies of responses within one question; pie charts are unable to represent more than one level of data. Bar and column graphs as well as line graphs are excellent ways to depict more complex information.

Making a Pie Graph

To illustrate how to produce a pie graph in Excel, we will use the frequencies calculated for Question #1 on the sample spreadsheet (figure 3-11). Click on the cells where you have calculated the frequencies for Question #1. They will be outlined in color. In the sample spreadsheet, this corresponds to the cells B13 through B16.

Now click on the “Insert” tab at the top of the screen (next to the “View” tab) and scroll to “Chart.” A variety of charts and graphs will appear just

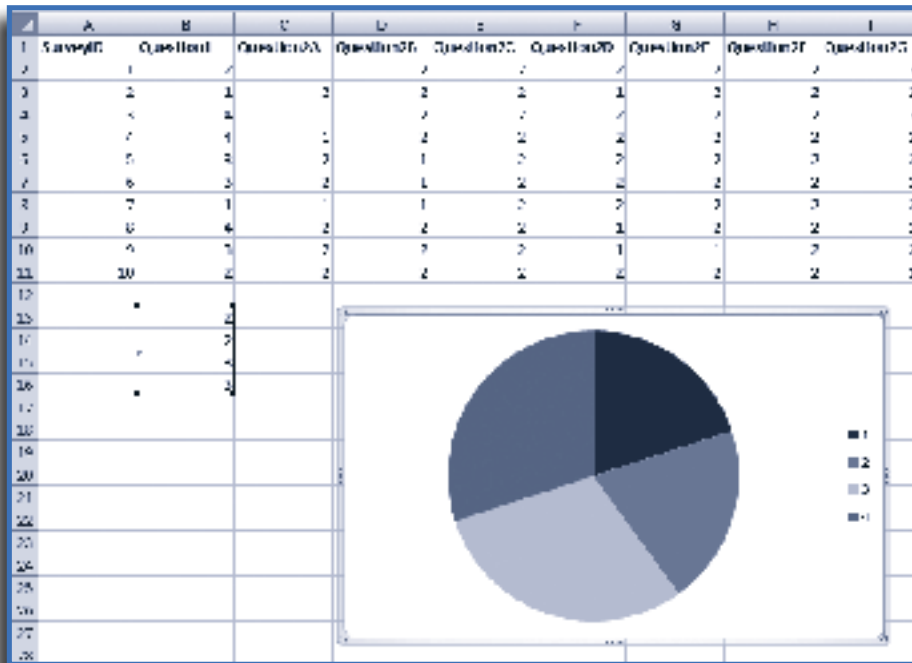


Figure 3-11: Sample spreadsheet illustrating data to be used in creating a pie graph.

above the spreadsheet. Click on “Pie.” Six graphic depictions of pie graphs to choose from will appear. Click on “Pie.” A graph with pie pieces should appear in the middle of the spreadsheet.

You can either click on the chart and drag it to a more convenient place on the spreadsheet or it can be copied and pasted into a report using Microsoft Word. Notice that the chart is not labeled at this point in order to describe to readers what they are seeing.

All charts and graphs must be adequately labeled with a title and a key describing what the relative values are. These features can be added by clicking on the chart and using the Layout tab under Chart Tools or the floating formatting palette to the right to add information and change the coloring, shading, and other qualities of the pie chart, such as shown in figure 3-12.

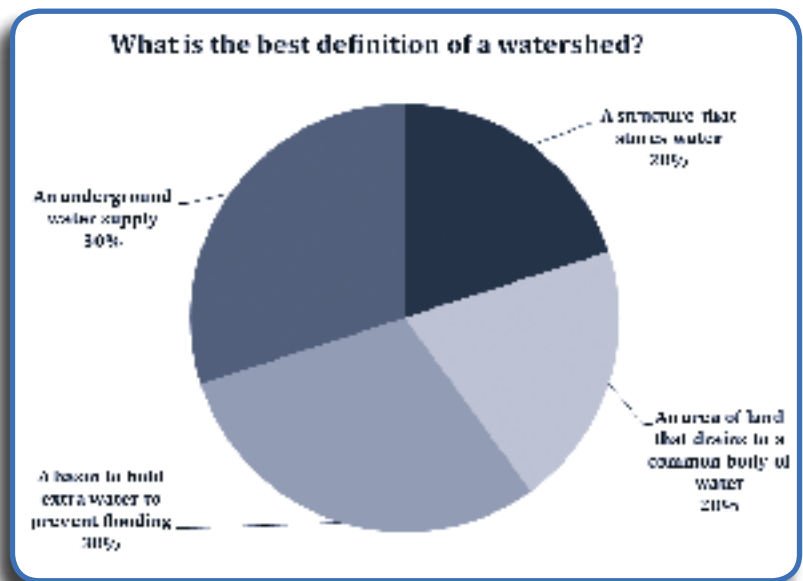


Figure 3-12: Sample pie graph with labeling and values added.

If you want to make a pie graph of the frequencies of responses for Question #2, take your cursor and select all of the “yes” replies (coded as “1” on the spreadsheet) across all the columns for Question #2 (labeled Question #2A through Question #2G). In the sample spreadsheet, these would be cells in Row 13 for Columns C through I. These cells should be highlighted in color as shown in figure 3-13.

	A	B	C	D	E	F	G	H	I
1	SurveyID	Question1	Question2A	Question2B	Question2C	Question2D	Question2E	Question2F	Question2G
2		1	2	1	2	2	2	2	2
3		2	1	2	2	2	1	2	2
4		3	4	1	2	2	2	2	2
5		4	4	1	2	2	2	2	2
6		5	1	2	1	2	2	2	2
7		6	1	2	1	2	2	2	2
8		7	1	1	1	2	2	2	2
9		8	4	2	2	2	1	2	2
10		9	3	2	2	2	1	1	2
11		10	2	2	2	2	2	2	1
12									
13									
14									
15									
16									
17									

Figure 3-13: Question #2 responses highlighted in order to create a pie graph.

Follow the same instructions as for the pie graph you made with Question #1 data. After cleaning up the graph and adding titles and percentages, the graph should look something like the pie graph in figure 3-14.

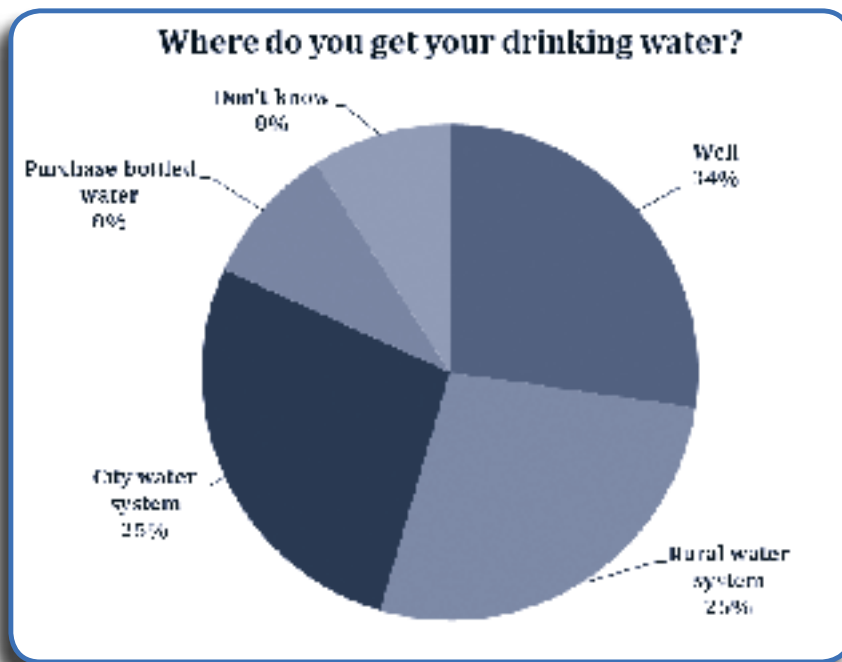


Figure 3-14: Question #2 answers illustrated in a pie graph.

Making a Bar or Column Graph

Bar and column graphs are essentially the same kind of representation of frequencies, except that a bar graph depicts the data horizontally and a column graph depicts the data vertically. Figure 3-15 uses the same data in a bar graph that was used to create the pie graph (figure 3-12). The pie graph shows data in percentages and the bar graph depicts the data in actual numbers. The numbers after each bar on the far right represent the number of survey participants that chose that response.

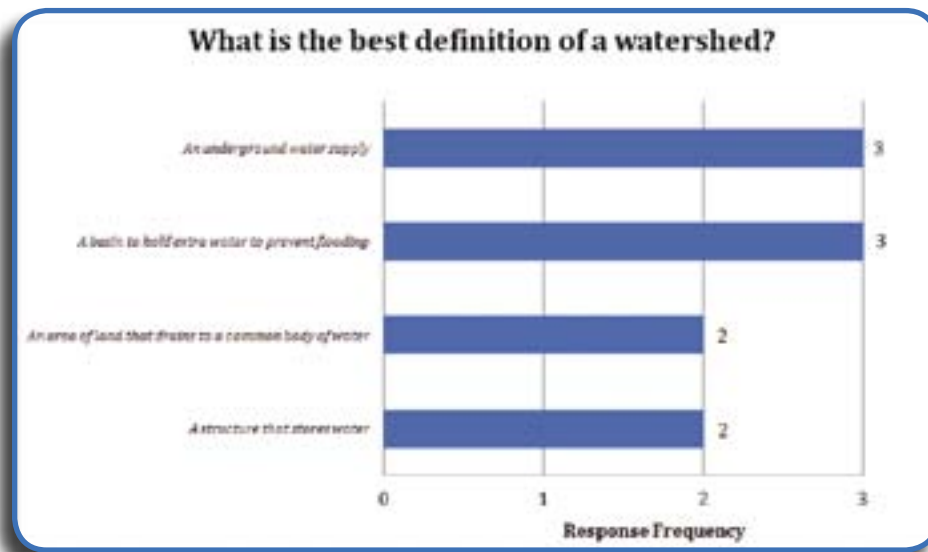


Figure 3-15: "Water Issues in Iowa" survey Question #1 responses shown in a bar graph.

While bar and column graphs can depict frequencies for simple questions like that of a pie graph, they are ideal for depicting more complex information with multiple variables. Say, for instance, you wanted to graph the responses to Question #6. Because pie graphs can only depict one level of data, you could only represent the data in each category found in Question #6 separately.

Because there are eight categories, that would mean eight separate pie graphs! Instead, you can represent all the data in Question #6 on the same bar or column graph. To do so, click with your cursor on the first set of frequencies for Question #6A in Cell M13 (Column M, Row 13).

Drag the cursor down to cover all four frequencies under Question #6A and then across until you have reached the column labeled Question #6H. This corresponds to the area between Rows 13 and 16 and Columns M through T. The area should be highlighted blue. Now click on the “Insert” tab at the top of the screen (next to the “View” tab) and scroll down to “Chart.”

A variety of charts and graphs will appear just above the spreadsheet. Click on “Column.” Several graphic depictions of column charts to choose from will appear. Click on the choice labeled “2-D Clustered Column.” A column chart looking something like the one in figure 3-16 should appear.

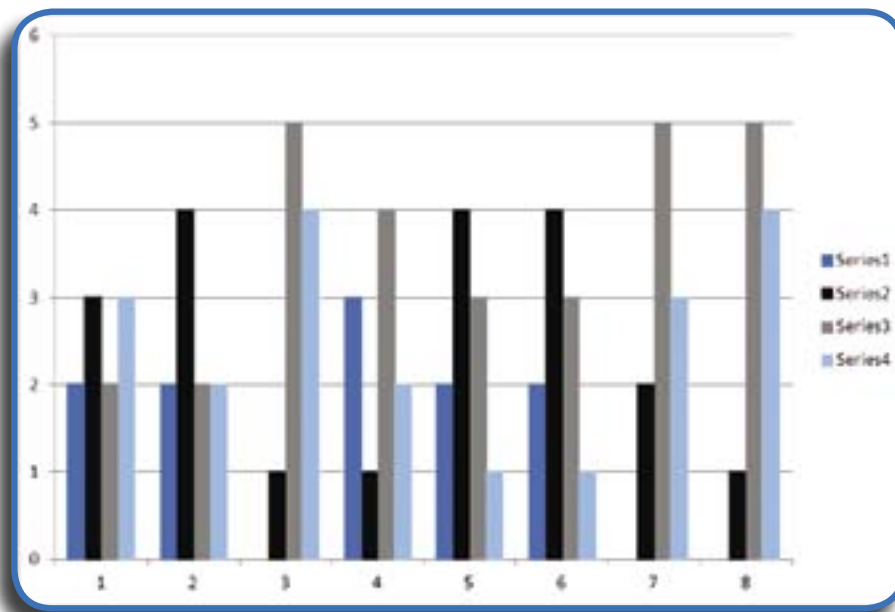


Figure 3-16: Bar graph showing the answers to Water Issues in Iowa survey Question #6 without labels and supplemental information.

You will have to label this appropriately so that the data is interpretable and understandable by others. Make sure to add a title to the graph. You will need to label each value along the horizontal portion (numbered 1 through 8). These numbers represent each of the conditions asked about on Question #6.

Each bar of the graph clustered above the horizontal number represents the frequency of each response (“know,” “suspect,” “not a problem,” and “don’t know”). In figure 3-16 the legend has these different responses labeled as “Series 1 through 4.” You should change these to represent our response categories. A cleaned up graph may look like the bar graph in figure 3-17.

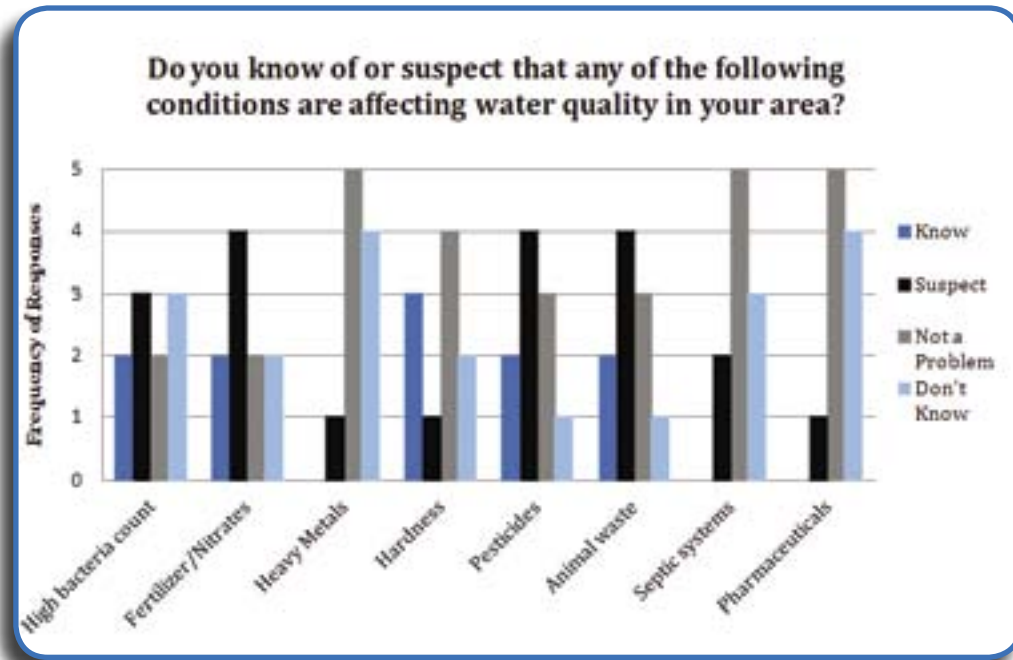


Figure 3-17: Bar graph with labels and title added.

Excel offers many options for different graphs. Some trial and error is recommended to become familiar with different graphing techniques and how they represent your data differently. It is important to keep in mind that not every type of graph is best suited to display the trends in your data. Make sure that the style of graph you choose represents the data in the way you want it to be interpreted.