



Graphical Representations

Introduction

Graphical representations are visual ways to represent raw data, descriptive statistics, or inferential statistics in order to highlight important findings from a research study. There are many different kinds of graphical representations; it is up to you to decide which best describes your data (Shaughnessy, Chance, and Kranendonk 2009). Then, you can use the graphical representations to determine additional statistical analysis that can be done to explain the significance of the data.

Learning Objectives

During the course of this chapter you should

1. familiarize yourself with the various graphical representations for quantitative and qualitative data,
2. determine which types of graphical representations should be used for your own data,
3. construct a variety of graphical representations in your laboratory notebook, and
4. make notes in your laboratory notebook about what each graphical representation highlights in your data.

Key Terms

Bar graph: Shows the distribution of qualitative data using bars that indicate frequency of a specific category.

Box and whisker plot: A graphical representation showing the distribution of quantitative data; the plot displays minimums, maximums, outliers, ranges, interquartile ranges, medians, and lower and upper quartiles.

Cartesian coordinate system: A system that plots points uniquely in a plane by a pair of numerical coordinates; the system indicates the distances from the point to two perpendicular lines (y -axis and x -axis).

Dot plot: A graphical representation showing the distribution of quantitative data using dots to show the relationship between two (or more) variables.

Frequency distribution: Shows the amount of variation within a spread of qualitative data that counts the frequency of data points of the values in the sample.

Graphical representation: A visual way to represent raw data, descriptive statistics, or inferential statistics in order to highlight important findings from a research study.

Histogram: A graphical representation showing the distribution of quantitative data using touching bins (see p. 115) that indicate frequency of a specific value.

Inferential statistics: Mathematical calculations performed to determine whether the differences between two groups are due to chance or are a result of the treatment.

Line graph: A type of graphical representation using the Cartesian coordinate system that displays data points connected by line segments; often used to show change in time, with time being on the x -axis.

Normal distribution: A curved shape in a graph that shows increased frequencies centered around a single mean and fewer frequencies on either side of the mean.

Pie chart: A circular chart that represents data by dividing the whole circle into sectors by sizes that are proportional to the quantity the sector represents.

Scatter plot: A type of graphical representation using the Cartesian coordinate system that displays both the independent variable and dependent variable as a single point; often used to determine positive correlation, negative correlation, or no correlation between two variables.

Stem and leaf plot: A graphical representation showing the distribution of quantitative data to visually highlight the frequency of certain values, without losing the original, or raw, data.

Table: A type of graphical representation that organizes numerical and/or descriptive data in rows and columns that aid in the understanding of a large amount of data.

Construction of graphical representations can be done manually or by using technology such as a calculator or a spreadsheet program; consider a graphing calculator such as the TI-Nspire (or others in the TI family) or statistics software such as Fathom Dynamic Data Software.



Whatever method of construction you choose, you should enter all graphical representations into your laboratory notebook under the laboratory heading Graphs, Statistical Analysis, and Conclusions. As first discussed in Chapter 6, “Organizing a Laboratory Notebook,” you should adhere, or link, to each table, graph, or statistical test and write down your observations and an analysis of what each graphical representation might show about the data. It is critically important that you label all graphical representations carefully so that you have an accurate record of exactly what data have been organized. By doing that, you will be able to accurately interpret the results in the days or even weeks after the data have been placed in your laboratory notebook.

Graphical Representations for Quantitative Data

The type of graphical representation you choose will be based on the type of data you have: quantitative or qualitative. This section introduces common types of graphical representations for displaying quantitative data: histograms, dot plots, stem and leaf plots, box and whisker plots, line graphs, scatter plots, and tables.

Histograms

A histogram is a graphical representation showing the distribution of data (McDonald 2009). At first glance, histograms may look just like what you have always called bar graphs. Although both types of graphs are constructed with horizontal or vertical bars, a histogram’s bars are called *bins*, and these bins touch, with no spaces between them. In addition, the numerical data represented by the axis at the base of the bins are on a numeric continuum, with frequency being indicated by the height of each bin. Also, histograms are best used for quantitative data; bar graphs are better used for qualitative data. For example, a histogram should be used if you need to construct a

Tips for Generating Graphical Representations (“Visuals”)

- Understand what it is you want to display
- Plan your visual before creating it
- Choose the best technology for the type of visual
- Communicate only one idea in a single visual
- Label everything to increase understanding
- Keep it simple; do not display too much information
- Maintain the scale by visually keeping equal distance between values on the x -axis and y -axis.

graph comparing the number of M&Ms in 23 different bags. But a bar graph would be used to compare the number of different color candies found in the bags of M&Ms.

Let's explore histograms further. Suppose the number of M&Ms in 23 separate bags are counted. Listed below are the raw data results with the amount of M&Ms in each bag of candy.

22, 23, 23, 23, 22, 24, 25, 21, 23, 24, 24, 23, 22, 21, 23, 23, 20, 24, 26, 22, 23, 22, 22

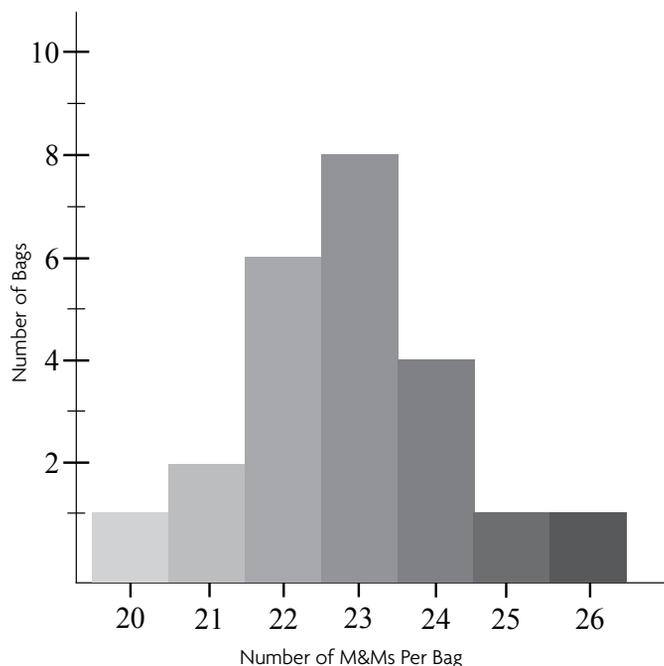
The numbers listed this way make it difficult to determine how many M&Ms are most commonly found, or least likely to be found, in a bag of M&Ms. Therefore, the first task is to arrange the data in order from least to greatest.

20, 21, 21, 22, 22, 22, 22, 22, 22, 23, 23, 23, 23, 23, 23, 23, 24, 24, 24, 24, 25, 26

These data are perfect for a histogram. If we make it using vertical bins, our horizontal axis will be labeled with the varying quantities of M&Ms. Our data is best displayed by constructing bins that are one unit wide, starting with 20 and going up to 26 (see Figure 8.1). The vertical axis will be labeled *frequency*, or the number of bags that contained that quantity. Therefore, frequency represents the numbers that fall within the interval of that bin. There is only one bag that had 20 pieces of candy, so the first bin is one high. There are two bags that had 21 pieces so the next bin is two high, and so forth.

Figure 8.1

Histogram Frequency of M&Ms Found in 23 Bags



GRAPHICAL REPRESENTATIONS

Figure 8.1 is an example of a *unimodal* histogram. This means there is one clear mode, or peak, in the data. This histogram can also be called roughly symmetric because both sides of the histogram look similar. Since this distribution has the largest bins at 22 and 23, it seems that 22 and 23 are the most common quantities of M&Ms that will appear in the bag. The bins are the smallest at the tails, or ends, of the histogram. It appears as if 20 and 26 are uncommon quantities of M&Ms. This curved shape that shows increased frequencies centered around a single mean and fewer frequencies on either side is known as a *normal distribution*. Because this sample of M&Ms has no bin that goes outside this curve, we can say the sample is normally distributed.

Constructing a histogram by hand is time-consuming and difficult to do well. To obtain professional quality, more efficient histograms can be constructed using either Texas Instruments software or calculators. The TI-84 and TI-Nspire families as well as Fathom statistical software can construct high-quality histograms.

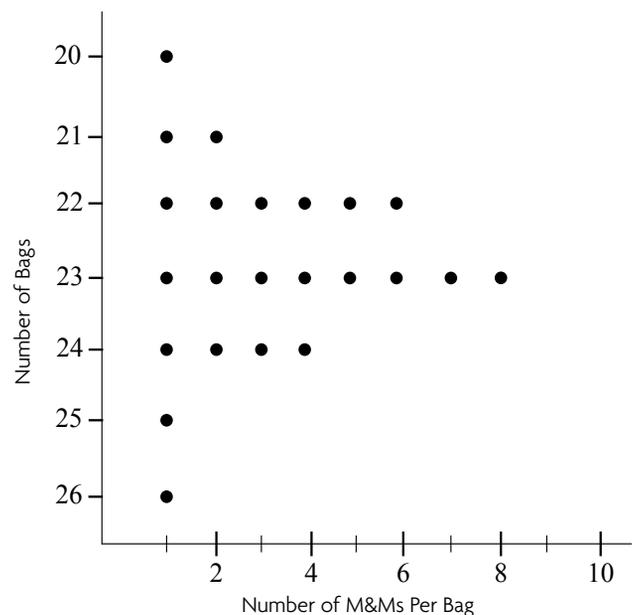


Dot Plots

A dot plot, also called a strip chart or strip plot, is another way to graphically represent the distribution of quantitative data, using dots instead of bins. Similar to the histogram, dot plots have one axis labeled like a number line and another axis labeled with frequency (the quantity of data that falls within a given interval). When a dot plot is constructed, it will look similar to a histogram. In fact, the overall shape of the dot plot will be exactly the same as if you constructed a histogram. For example, consider the histogram of the M&Ms we just completed. Instead of making bins for each interval, we can place a dot for each M&M bag that falls into this category, as shown in Figure 8.2. Data within both dot plots and histograms can be oriented either vertically or horizontally. The dot plot in Figure 8.2 is shown horizontally. A dot plot can be constructed with the TI-Nspire and with Fathom statistics software.

Figure 8.2

Dot Plot Frequency of M&Ms Found in 23 Bags



Stem and Leaf Plots

A stem and leaf plot is another way to display quantitative data. It is best used when you want to visually highlight the frequency of certain values, without losing the original, or raw, data. The distribution of this data will look similar to both a histogram and a dot plot, except the stem and leaf plot shows the numeric value of *each individual piece* of data in addition to grouping them as a whole. To organize your data, you will separate your data into both a stem and a leaf. The first column is called the stem and the second column is the leaf. The list of 20 values below represents weights in grams.

46, 34, 59, 44, 45, 55, 38, 47, 55, 31, 44, 56, 33, 45, 30, 58, 46, 49, 40, 48

Although not absolutely necessary, ordering the list makes it easier to plot.

30, 31, 33, 34, 38, 40, 44, 44, 45, 45, 46, 46, 47, 48, 49, 55, 55, 56, 58, 59

The first column, or stem, for our example will contain the tens digits while the second column, or leaves, will contain the ones digits. Therefore, the stem-leaf plot would look like Table 8.1.

Table 8.1

Stem and Leaf Plot of Weight in Grams

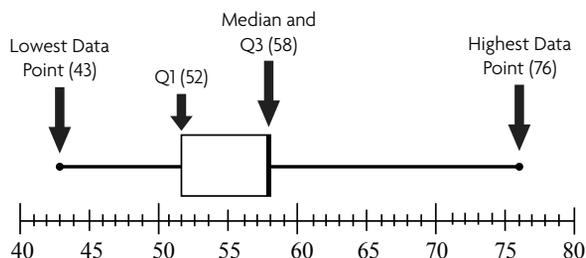
Stem (g)	Leaf (g)
3	0 1 3 4 8
4	0 4 4 5 5 6 6 7 8 9
5	5 5 6 8 9

Notice that the horizontal leaves in the stem and leaf plot would correspond to the vertical bars in a histogram showing the same data, and the leaves have lengths that equal the numbers in the frequency table. From this type of graphical representation, it is obvious at a glance that there were more data in the 40s than in the 30s or 50s. However, the raw data values can still be determined. For example the bottom leaf shows that the five values in the 50s were 55, 55, 56, 58, and 59.



If you have a lot of data to enter into a stem and leaf plot, preparing it manually may introduce researcher error. You must still pay attention to details when using technology to make sure you don't make any mistakes, but using technology may save you time. If you use a spreadsheet program, you can enter your data into a column, highlight the column, and then sort the data in ascending order before you begin the construction of your stem and leaf plot.

Figure 8.4

Box and Whisker Plot for Distances a Ball Traveled on Concrete

The box and whisker plot shown in Figure 8.4 is a horizontal version of a box and whisker plot, but it could have also been created vertically. The benefit of this type of graphical representation is that it concisely displays a lot of data. It shows the range of distances the ball traveled, with the shortest distance being 43 cm and the longest being 76 cm. The plot also shows that the median—which also happens to be the same number as Q3—is 58. That indicates that 25% of the numbers are this value. Another 25% of the data numbers fall between 52 and 58. The lines, or whiskers, indicate that the rest of the data numbers fall between 43–52 and 58–76.

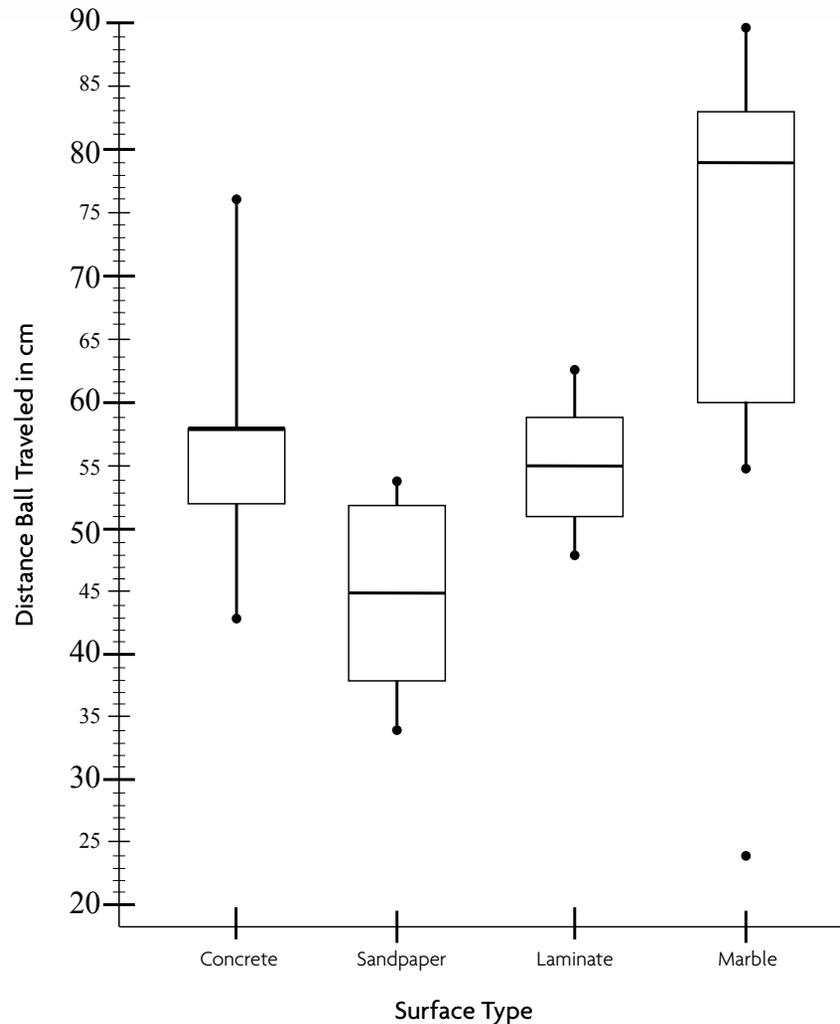
For the same experiment, Table 8.2 shows the results of the trials completed on additional surfaces and can be used to create a box and whisker plot, as shown in Figure 8.5.

Table 8.2

Distances the Ball Traveled on Different Surfaces

Surface	Distance Ball Traveled (cm)
concrete	43, 49, 52, 53, 56, 58, 58, 58, 58, 63, 76
60-grit sandpaper	34, 36, 38, 39, 41, 45, 50, 52, 52, 52, 54
laminated	48, 51, 51, 52, 53, 55, 56, 57, 59, 60, 63
marble	24, 55, 60, 75, 79, 79, 80, 82, 82, 85, 90

Figure 8.5

Box and Whisker Plot—All Surfaces

Box and whisker plots display data so that the groups can easily be compared. The axis in Figure 8.5 is oriented vertically, but it could also be constructed horizontally. The same steps are used to plot several sets of data together on a single graph that were used for plotting a single set. Notice that the shortest distance the ball traveled on the marble surface, 24 cm, although plotted, is not connected to the whisker. That is because, mathematically, it was determined to be an outlier. See the section “Dealing With Outlier Data” in Chapter 7 on pages 102-103.

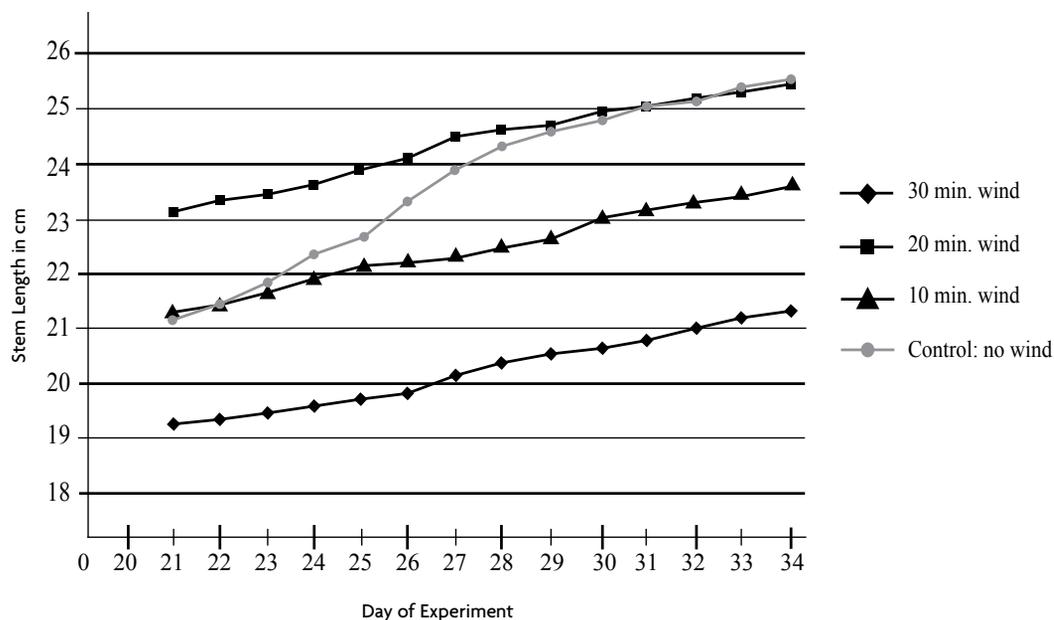
Line Graphs

A line graph displays data points connected by line segments. A line graph is often used to show change in time. The Cartesian coordinate system allows you to plot points uniquely in a plane using a pair of numerical coordinates. This system displays the distances from the point to two perpendicular lines (y -axis and x -axis). You may remember this concept from algebra and geometry classes. Consider using line graphs if the measurements of an entity's change over the experiment are significant or there are drastic differences between the groups. Time is plotted on the horizontal axis or x axis, with your dependent variable measurements on the vertical axis or y axis. By using different types or colors or patterns of lines, you can compare entities of several groups at one time.

For example, Figure 8.6 shows the effect of varying amounts of wind on the length of plant stems. The time, or days of the experiment, appears on the horizontal axis, while the dependant variable of stem length appears on the vertical axis. All the experimental groups, as well as the control, can be placed on the same graph.

Figure 8.6

How Plant Stem Length Is Affected by Various Amounts of Time in the Wind



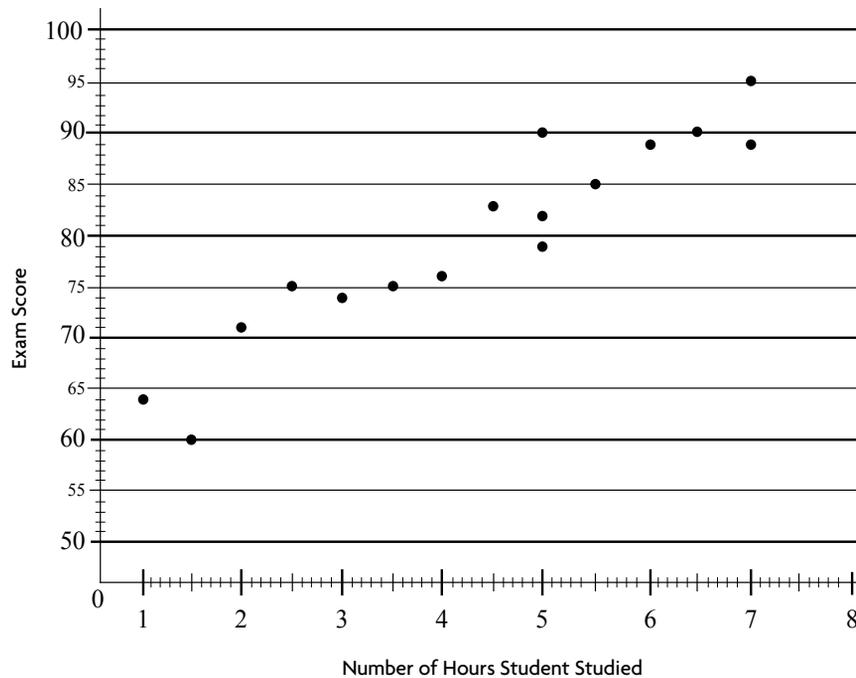
GRAPHICAL REPRESENTATIONS

Use line graphs cautiously because they can be misleading. For example, the control group in Figure 8.6, which had no exposure to wind, grew the longest stems during the experiment represented by the steep line. However, someone might look at the graph and, noting that the 20 minutes line is the highest on the graph, might interpret this group as having grown the longest stems. Therefore, while line graphs do help show trends over time, as well as plot raw data, they are not always the best choice for graphical representation. In some cases, tables that indicate total change, or percent change, more accurately represent the data. Another misconception can arise if you make an assumption regarding the data between the data points collected and plotted in your graph. The only real data are the dotted points. It is possible that by connecting the points, you are misleading your readers. You are making an assumption—and, therefore, a prediction—about the points between the actual plotted data points. You can address this assumption when you display the data in the Results section of the final paper or poster.

Scatter Plots

Scatter plots, also known as scattergraphs, are statistical charts that plot the values of both the independent variable and dependent variable as a single point (Triola 2001). A scatter plot graphically represents a pair of coordinates (x, y) on the Cartesian plane. Scatter plots are visually similar to line graphs and help display whether or not there is a correlation between the independent and dependent variable. The independent variable is customarily plotted on the horizontal axis and shows what was measured; the dependent variable is plotted on the vertical axis. If the pattern of dots slopes from lower left to upper right, it suggests a positive correlation, and if the dots slope from upper left to lower right, it suggests a negative correlation. Patterns of dots can also indicate nonlinear correlations. The scatter plot example in Figure 8.7 is showing a correlation between the number of hours a student studies and his or her exam scores.

Figure 8.7

Positive Correlation Between Exam Scores and Number of Hours Studied

The most noticeable difference between scatter plots and line graphs is that in scatter plots the dots are not connected with a line. Notice in Figure 8.7 that for five hours, more than one dot was plotted. This indicates that three students studied five hours and each of their scores was plotted. This is an advantage of using a scatter plot over a line graph—research that has several data points for a single interval can each be plotted. Also notice that the dots are generally sloping to show a positive correlation. The closer the dots are to an imaginary line, the stronger the correlation: the more spread out the dots, the weaker the correlation.

Line graphs and scatter plots can be constructed using any graphing calculator or in spreadsheet programs using the function called XY (Scatter).

**Tables**

A table organizes numerical and/or descriptive data in rows and columns that aid in the understanding of a large amount of data (Cothron, Giese, and Rezba 2006). Tables are often overlooked when students begin organizing data for visual display. While histograms, box and whisker plots, line graphs, and scatter plots are fun to construct, do not underestimate the ability of a table to

GRAPHICAL REPRESENTATIONS

clearly display data. It is good practice to provide raw data in the presentation of your results (Gordon 2007). This can be done in graphs, but tables are another good way to organize this information. Table 8.3 shows a summary of the number of egg masses that were counted at various places in a pond as well as the number of confirmed fertilized eggs for each of the pond locations. Having the raw data organized in a table is helpful to readers when you begin describing the data in the Results section of your paper, poster, or presentation.

Table 8.3**Summary of Leopard Frog Egg Mass Pond Observations**

Pond Identification Number	Number of Confirmed Egg Masses	Confirmation of Fertilized Eggs
38-VP-2	8	Yes
40-VP-2	31	Yes
19-VP-5	2	No
23-VP-5	0	N/A

Once raw data have been calculated via descriptive and inferential statistics, tables are helpful to compare the groups. Tables constructed with data from each of the groups help to highlight which groups had the most change and how groups compare to one another. For example, in Table 8.4 it is clear that as the level of hormone increased, so did the root growth.

Table 8.4**Total Root Growth and Rate of Growth of Plants Exposed to Different Levels of the Hormone Cytokinin**

Plant	Independent Variable	Average of Total Change in Root Length Growth (in mm)	Total Time (in days)	Average Rate of Growth (in mm/day)
Control Group Plants #1–4	No hormone	8	20	0.4
Experimental Group 1 Plants #5–8	2 ml of cytokinin	14	20	0.7
Experimental Group 2 Plants #9–12	4 ml of cytokinin	22	20	1.1

Notice in Table 8.4 that groups with names like “Experimental Group 1” are also identified by how the independent variable was manipulated.

You may feel as if you keep repeating the same information in the figures, tables, and narrative. This is good! The repetition is necessary. You don't want your reader to have to flip back to an earlier page to see what each group number represents.

Tables can be made in word-processing programs or in spreadsheet programs. The data in tables and graphs must be organized in such a way that the tables and graphs can “stand on their own”—that is, no additional narrative explanation is needed. You can accomplish this by labeling rows and columns with complete descriptions. However, please note that in your research paper, you *must* mention all graphical representation of data in the narrative.

Graphical Representations for Qualitative Data

In addition to your quantitative data, you may also have collected qualitative data. Qualitative data can sometimes be graphically represented in bar graphs, pie graphs, or tables. However, it's likely that much of your qualitative data will be described in the narrative of your paper. The key is to organize the data so you can analyze the correlation between the independent and dependent variable, taking into account the qualitative variables.

Bar Graphs

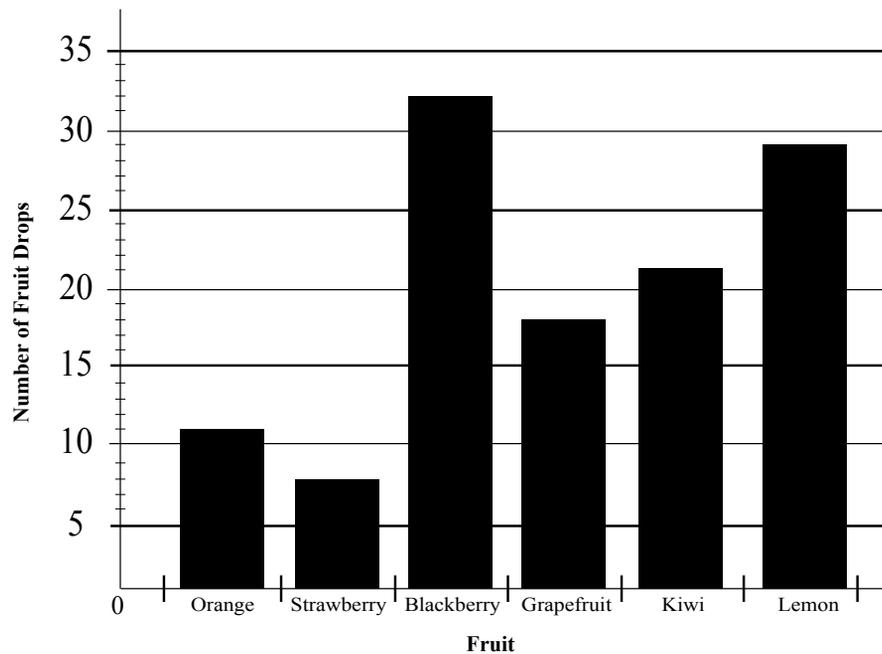
Bar graphs show the distribution of qualitative data using bars that indicate frequency of a specific category. Bar graphs look similar to histograms but are best used for qualitative independent variables. For example, if you designed a categorical system to record the odor of a chemical reaction, each of the experimental groups and the control would be bars on the horizontal axis, and the vertical axis would be an odor rating of descriptions, such as *very strong*, *strong*, *noticeable*, *weak*, or *very weak*.

The data in Figure 8.8 is from an experiment where 50 drops of the blue reagent, Vitamin C Indicator, or dichlorophenolindophenol, were placed into test tubes. Drops of various fruit juice were added one at a time, until the blue solution turned clear. Higher numbers of fruit juice drops indicate lower levels of vitamin C. This data belongs in a bar graph, not a histogram, because the order in which the fruit juices are listed on the horizontal axis does not matter. This is because the types of fruit juice are independent of one another

and are not on an incremental scale, which is how histograms are constructed. Notice, as well, that the bars in the graph do not touch.

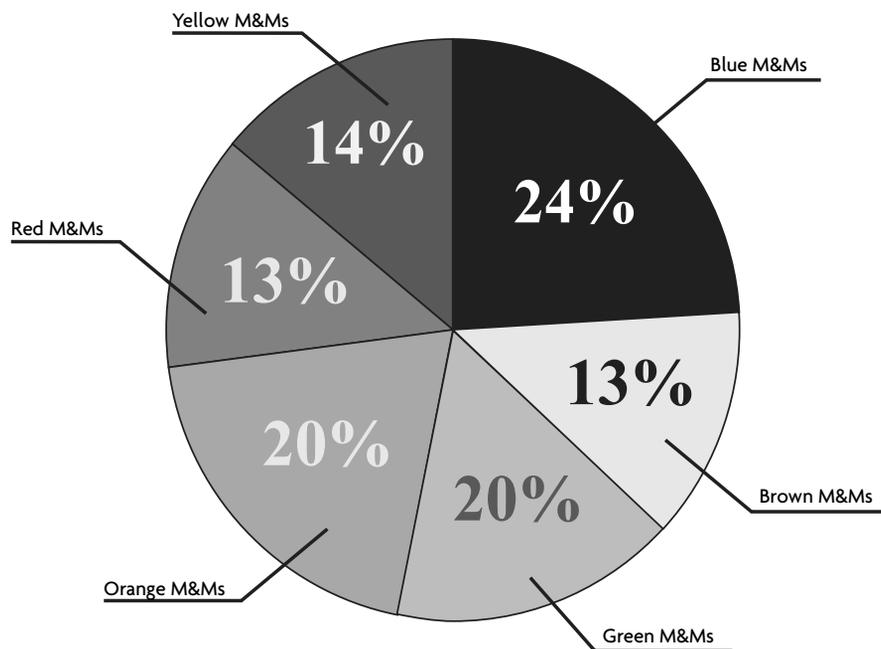
Figure 8.8

Number of Drops of Fruit Juice It Takes to Turn Vitamin C Indicator Clear



Pie Chart (or Circle Graphs)

A pie chart is a circular chart that represents data by dividing the whole circle into sectors by sizes that are proportional to the quantity the data represent. Pie charts are best for showing how data pieces fit together as a whole but are less effective for comparing data to one another. A full circle represents a whole, or 100%, and the sectors, or slices within the circle, represent the parts that make it up. For example, Figure 8.9 shows the average percent of the varying colors of M&Ms within a single bag. We understand that the circle represents a full bag, and each slice represents the percentage of a particular color in the bag. Although a pie graph is visually appealing and easy to construct in spreadsheet programs, in the case of Figure 8.9 a data table might better compare this particular set of data because each of the pie sizes is similar in size.

Figure 8.9**Average Percentage of Different Colors of M&Ms in a Single Bag****Tables and Narrative**

Tables work well for qualitative as well as quantitative data. In your laboratory notebook, review your written record of procedures as well as any data tables about which you wrote a narrative. The descriptive observations you wrote on the entities themselves throughout the experiment may highlight something not captured by your quantitative data. Carefully review these narratives for patterns and trends as well as outlying and strange occurrences. This is also the time to review the photographs you took throughout the experiment. Do you notice anything different from what you recorded as quantitative data? While typing out descriptions of each entity throughout the experiment is not recommended, a table describing a condensed version of these observations would be helpful if there are noticeable trends or strange occurrences. Table 8.5 shows what a sample of what a qualitative table might look like within a research paper. Even if you decide not to insert a table such as this in your research paper, the observations you made should be discussed when you write your paper to extract additional meaning from the quantitative data.

Table 8.5

Summary of Qualitative Data and Research Journal Entries

Experiment Day	Qualitative Data and Research Journal Entries
5	After five days, there was still no mold growth on any of the bread slices. Therefore, the experimental setup was altered. To increase humidity, each bread slice was placed inside a gallon-size resealable plastic bag with a paper towel saturated with 75 ml of water.
10	Although no mold was growing on the bread slices themselves, there was a red splotch on one of the paper towels. Viewing the splotch using a magnifying glass and microscope, I found fibrous extensions of the red mass, which helped to determine that the red mass was biological, not chemical.

Making Final Graphical Representations

Using Graphical Representations to Explain the Data

The purpose of organizing your data into graphical representations is to provide a new perspective of the data. A visual representation will help you better determine what differences, if any, exist between your experimental and control groups. As you look for trends, patterns, and differences highlighted by your graphs, ask yourself questions about what you see. (See sample questions on pp. 140–142, in the “Data Interpretation” section.)

Like descriptive statistics, however, the differences highlighted by your graphs may not necessarily be the result of the changes you made to the independent variable. It is possible that these changes are due to experimental error or to chance. Therefore, inferential statistics should be calculated on the differences you notice in your graphical representations. See Chapter 9 to learn more about inferential statistics.

Reviewing the Records of Procedures and Correspondence

Another important part of organizing data is to reread your laboratory notebook sections titled “Record of Procedures” and “Record of Correspondence.” This review may highlight how procedures or events affected the outcome of your results. Start by rereading your record of procedures with a highlighter and mark anything out of the ordinary. “Out of the ordinary” might be a weekend you were gone and your sister forgot to check the experiment or when someone dropped group #2 and the experimental setup had to be reconstructed. Take time to compare the dates of these unusual events with the actual quantitative data. Are there correlations between the changes in the data and the time of the event, or shortly after? If not, within the Analysis and

Conclusions section of your paper or poster, you will be able to say that these occurrences did not appear to affect the results. However, if there was a correlation, you will need to explain how and why those occurrences may have influenced the results. In addition, determine if the Record of Procedures events explains any quantitative outlier data. If so, you will need to defend your decision to keep or remove the data from the analysis. Refer to the section titled “Dealing With Outlier Data” in Chapter 7 for more information. Record your connections and inferences in the “Graphs, Statistical Analysis, and Conclusions” section of your laboratory notebook.

Consider using tables to organize any of the entries that may be important. Although you should not create a table that includes the entire Records of Procedure and Correspondence, you may want to construct a table that shows only the modifications to the proposed methods and/or occurrences that may have affected the data.

Preparing Graphical Representations for the Paper and Poster

When preparing graphical representations to include in your results and analysis and conclusions sections of your paper and/or poster, there are a few important details to keep in mind. Labels within figures and tables are extremely important. Use experiment-specific labels when describing axes and other important features within the figure and columns and rows within tables. For example, provide a description of the independent variable treatment. In Figure 8.6, the legend clearly defines each line by how the independent variable was manipulated: 30 min. of wind, 20 min. of wind, and so forth. The legend should *not* be labeled generically as “Experimental Group 1.” If it were, the graph cannot be interpreted without having to look back into the narrative to see what treatments each of the groups received. Figure and table titles are also important. Each must be specific for the data it portrays. Labels must clearly describe what is contained in the graphical representation, and the horizontal and vertical axes and table columns and rows should be labeled clearly with descriptive words and units. See Chapter 11 for additional detail on how to format entire figures and tables for a paper or poster.

Chapter Applications

In your laboratory notebook, you should construct, and comment on, many different graphical representations of your data. Remember to use informative labels and detailed titles for each table or graph. Most important, make comments near the visual about your observations and your analyses of what the visual may indicate about your data, about the experiment as a whole, and specifically about the relationship between the two variables.

References

- Cothron, J. H., R. N. Giese, and R. J. Rezba. 2006. *Science experiments and projects for students: Student version of students and research*. Dubuque, IA: Kendall/Hunt.
- Gordon, J. C. 2007. *Planning research: A concise guide for the environmental and natural resource sciences*. New Haven: Yale University Press.
- McDonald, J. H. 2009. *Handbook of biological statistics*. Baltimore: Sparky House Publishing.
- Shaughnessy, J. M., B. Chance, and H. Kranendonk. 2009. *Focus in high school mathematics: Reasoning and sense making in statistics and probability*. Reston, VA: National Council of Teachers of Mathematics.
- Triola, M. F. 2001. *Elementary statistics*. New York: Addison-Wesley.
- Valiela, I. 2001. *Doing science: Design, analysis, and communication of scientific research*. New York: Oxford University Press.