Data Analysis

Collecting and analyzing data are important steps in the scientific process. We need to analyze data to see how the data support (or do not support) our hypotheses.

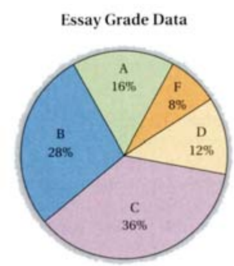
Graphing is one method used to help us analyze data we collect. A graph is a picture that shows a relationship between two quantities. A graph can transform data in ways that demonstrate aspects of the data more clearly and meaningfully such as showing patterns and trends or summarizing information.

Good graphs explain the results of the exercise and help us pose questions and guide us to new or revised hypotheses. In order to graph well, you must be able to choose appropriate graphs for particular data sets and you need to be familiar with the language of graphing

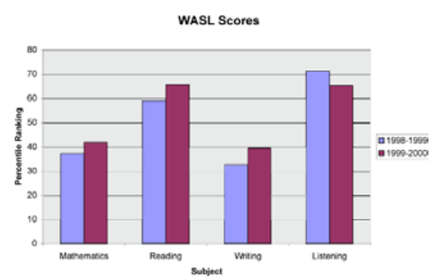
- Data
  Data are pieces of numerical information. Data can be displayed in tables, in graphs, or as sets of ordered pairs. (Data is actually plural. The singular form of data is datum.)

- Graph
  A graph is any visual representation of a data set. It may be a pie chart, a bar graph, a line graph, or a scatterplot.

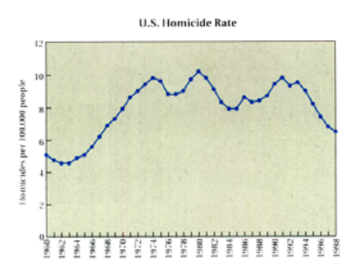
  - Pie chart
    A pie chart is a circular graph, cut into wedges or slices. Pie charts are used primarily for data that adds to 100%. The size of each pie slice is proportional to the percentage of the category it represents.



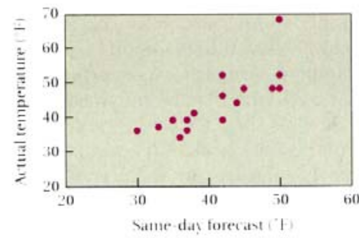
  - Bar graph
    Bar graphs display quantities (data) as bars that run either vertically or horizontally. The bar length corresponds to its numerical value.

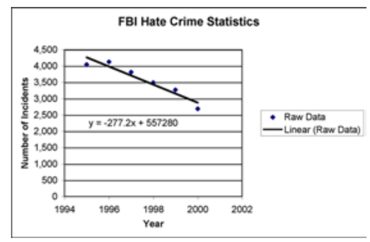


  - Line graph
    In a line graph each point represents one datum and consecutive points are connected by line segments. A line graph is often used to show how some quantity changes with time.



- Scatterplot
  A scatterplot shows the relationship between two numerical variables measured on the same individuals. The values of one variable appear on the horizontal axis, and the values of the other variable appear on the vertical axis. Each individual in the data appears as a point in the scatterplot.





- Best fit, or regression line
  A regression line is a straight line that describes how a dependent variable, y, changes as an independent variable, x, changes. We often use a regression line to predict the value of y for a given value of x.

- Variables
  The two quantities in a relationship are called variables because they change or vary. For example, in a relationship between time and temperature, both time and temperature vary. Time steadily advances while temperature can rise and fall. In a time/temperature relationship, time is written first because it makes more sense to say that the temperature depends on the time than to say that time depends on temperature; time advances no matter what else happens. Therefore time has independence and is called the independent variable. Temperature is called the dependent variable because the temperature depends on time. When making a graph, the independent variable is plotted on the horizontal, or x-axis, and the dependent variable is plotted on the vertical, or y-axis. Although it's easy to "see" in the time/temperature relationship which is the independent and which is the dependent variable, the distinction is not always clear-cut. In complex systems there can be multiple variables that can be closely related and constantly changing. (Think about the variables in the US economy). In science we want to be able to control as many variables as possible to focus on the variable we are investigating.

- Slope
  Slope is a ratio of a change in the dependent variable to a change in the independent variable. In biology, slope is often called "rate of change". To calculate the slope between two points (x_1, y_1) and (x_2, y_2), find the difference between the y coordinates of the two points (this is "rise") and the difference between the x coordinates (this is "run"). We usually use the letter m to represent slope.

$$\text{slope} = \text{m} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in y}}{\text{change in x}} = \frac{y_2 - y_1}{x_2 - x_1}$$

  Slope can be written as a fraction or as a decimal, and it can be positive, negative, or even zero. The units of slope are the units of the y-variable divided by the units of the x-variable. (For example, if the x-axis is labeled with days, and the y-axis is labeled with inches, the units of slope in this relationship is *inches per day.*) Slope is meaningful because it gives us a relationship between the changes in the two variables. It is the familiar m in the slope-intercept form of a linear equation:

$$y = mx + b.$$

Group Assignment – Graphing Practice

Data for the percentage of ABO blood types among United States population cohorts is shown below. Working in groups of no more than three students, make a bar graph, pie chart and line graph of the data.

ABO Blood Types in the United States Population

ABO Blood Type	% of US Population		
	White	Black	Asian
A	40	27	28
B	11	20	27
AB	4	4	5
O	45	49	40

Discuss the strengths and weaknesses of each of the three graphs.

Which graph represents the data most effectively? _________________. Discuss why this graph is more effective than the other graphs you made.

What type(s) of data can be displayed more effectively with the types of graphs that you would not use for the blood type data?

Evaluate your three graphs.
Are the graphs you made excellent, good, adequate, poor? ________________.

Identify criteria you used to evaluate your graphs.

Materials for this exercise were developed at the MAC (Mathematics Across the Curriculum) 2002 Summer Institute by Jennifer Laveglia and Rosemary Richardson, Bellevue Community College, liberally assisted by Jennifer McFarland, Edmonds Community College.