

Notice of conditions of use

The following materials were developed as a supplement to the Elementary Quantitative literacy program at the University of Washington, in conjunction with workshops given to teachers of grades K-5 from the Seattle School district from 1998-2003. These materials are not intended to be a stand-alone description of graphs appropriate for use in elementary school. The document is dynamic, in that as we find additional issues that need clarification, the document is modified accordingly.

These materials may be used by others, as long as (1) they are not published or sold, (2) this notice of conditions of use is retained with the document, and (3) this document is referenced in publications that make use of the ideas.

The URL is: http://faculty.washington.edu/wijsman/wa_eq1.html

Ellen M. Wijsman
Division of Medical Genetics
Department of Biostatistics
BOX 357720
University of Washington
Seattle, WA 98195

Graphing styles: Kinds of graphs, when to use them, and how to prepare to use them for classroom use

The following pages are intended as a guide to different styles of graphing. The seven types of graphs and charts described, along with bar graphs which are already familiar to teachers, are all fundamental to data analysis. These eight types of data display encompass those described in both the state of Washington and the NCTM standards for mathematics for elementary school students. While not all types of data display will be appropriate for the very youngest children (K-1 or K-2), all should be useable for at least some data analyses in most third grade classes on up, with initial introduction of some concepts for these more sophisticated types of graphs in the later part of grade 2.

In order to prepare students for the assessments currently under development, it is important that they gain experience in using a wide variety of age-appropriate techniques of data analysis. The information included in the following pages consists of descriptions and examples of various kinds of graphs, instructions for preparing to use these plots in your classroom, and a brief discussion of when you might use one type of graph vs. another.

The seven types of data display described here are:

- Line plots
- Histograms
- Stem and leaf plots
- Two-way tables
- Box plots
- Scatterplots
- Time-series plots

In helping students to develop their data analysis skills, it is important to keep the following three points in mind:

- ⇒ For most data sets, there is more than one appropriate and informative way to plot the data.
- ⇒ It is valuable to see the same data plotted in several ways, both for the purpose of learning more about the data, and because this will help students to gain experience in choosing appropriate methods of data display and analysis. Professionals routinely do this in order to explore patterns in their data.
- ⇒ The graphing styles introduced here and incorporated into the kit activities are only a small fraction of those that could be used. Creative insights are an asset to good graphing and data analysis. If a student invents a different kind of graph that successfully makes a point about the data that can be interpreted by someone else, then the graph is useful.

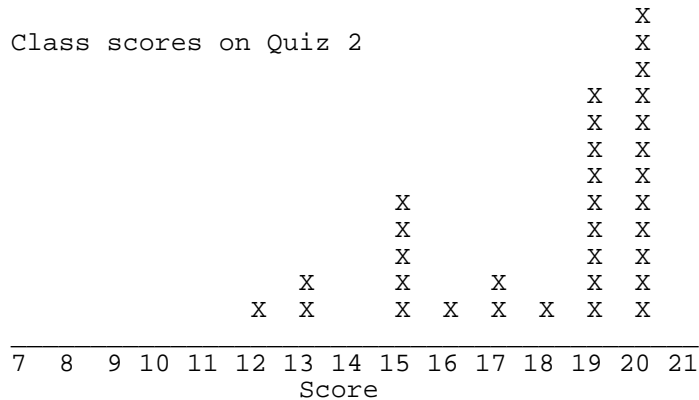
should cover the whole range of collected numbers, although for older children, you can “fudge” a little if there is one number which extends far beyond the range of the rest of the numbers (an **outlier**) by just plotting that point way out beyond the graph. For older children, use the actual range of the data; for the very youngest children (grade K, and possibly grade 1, especially early in the year), it is best to start with 0 or 1. Mark off the integers with lines on the tape—if possible, use 2” tape so you can make the numbers bigger. Place a title on the graph.

Graphing:

If children have not done this kind of plot before as a class, it is useful to model the process. Have one volunteer bring his/her post-it(s) up and step him/her through the process of where to put it. Repeat with a second student. Stress neatness (line up the post-its or dots up carefully so that they next to each other). Post-its have the advantage of being easily moved if they don’t get put in the right place. Also, in terms of data collection, even the smallest children can write their numbers on a post-it while dots may be too small. The final graph should look something like the quiz graph above, with the Xs replaced by dots or post-its.

Variations and Extensions:

Comparing data: In scientific investigation, we often wish to compare the results from two (or more) situations. If we have exactly two situations to compare where each could be represented on a line plot, a **back-to-back line plot** provides a convenient way to display the data. Here we plot *two* different situations on opposite sides of the number line describing the dependent variable. We must remember to label the two conditions on the graph! For the quiz example, we might have the data on quiz number 2:



Stem-and-leaf plots

Description:

A **stem-and-leaf** plot provides an alternative to a line plot or histogram for obtaining a picture of the data distribution when data can be represented as integer (counting) numbers. The resulting graph looks very much like a line plot or a histogram. However, this type of plot can be constructed very quickly—it can be used, for example, when a teacher wants to construct a graph on a chalkboard or an overhead using data collected by the children in the class. The children can simply call out their numbers one by one and the teacher can enter the numbers onto the graph.

An example data set:

Eric's third period PE class just had a fitness exam. Each child did as many sit-ups as he/she could do in one minute. Eric's score was 45. How well did he do relative to the rest of the class? The data for the class are as follows:

49	58	57	43	46	35	42	56
47	45	45	43	51	36	41	50
34	42	46	47	64	40	38	41
50	48						

A stem-and-leaf plot provides a convenient way to display the distribution of these data. In a stem-and-leaf graph, we separate the digits of each data point into the right most digit vs. the rest, so, for example, the 1s place becomes the “leaf,” and the rest of the number, becomes the “stem.” So in the above data set, we can split the numbers into the number of 10s (the stem) and the number of 1s (the leaf). 49 thus becomes $40 + 9$ or 4 tens plus 9 ones. We are not restricted to 2-digit numbers—for example, we could have collected forearm lengths in mm, in which case the numbers above might be 100 greater, so $149 = 14$ tens plus 9 ones, etc.

In constructing the stem-and-leaf plot we end up with something that looks very much like a line plot or histogram. There are two differences, however. First, we pool counts within a class of numbers, e.g. all counts 30-39, 40-49, etc. Second, we retain the actual numbers to make the graph, rather than using dots or Xs, so that we can continue to use the graph for determining other summaries of the data, such as the median. (Note that we lose this information when we use a histogram, although we can still obtain such summaries from a line plot).

⇒ Note: to get a good picture of the data, the final plot must be made quite neatly. To encourage this, it is often convenient to use graph paper, and to write each digit in a separate box on the graph paper.

The method of construction:

Step #1 Find the largest and smallest scores (the largest is 64 and the smallest is 34). If one person in the class is making the plot, this is easy to do orally by asking for a small/large number and then asking for successively smaller/larger numbers until there are no more extreme values.

Step #2 Because the smallest number, 34, has a 3 in the tens place, and the largest number, 64, has a 6 in the tens place, the stems will be the digits 3 to 6. For younger children you may want to simply write down all the tens place digits (1-9). Write these digits vertically with a line to the right.

```

Stem
3 |
4 |
5 |
6 |
    
```

Step #3 Separate each score into a stem (number of tens) and a leaf (the 1s digit). Write the leaf on the plot next to the stem. The first score is 49. The 9 is placed next to the 4 in the stem:

```

Stem Leaf
3 |
4 | 9
5 |
6 |
    
```

Continue until all the scores (from left to right) in the above data set are listed on the plot:

```

3 | 5 6 4 8
4 | 9 3 6 2 7 5 5 3 1 2 6 7 0 1 8
5 | 8 7 6 1 0 0
6 | 4
    
```

Step #4 On a new plot, rearrange the leaves so that they are ordered from smallest to largest.

```

3 | 4 5 6 8
4 | 0 1 1 2 2 3 3 5 5 6 6 7 7 8 9
5 | 0 0 1 6 7 8
6 | 4
    
```

Step #5 Add a title and a key to the plot.

```

                Sit-ups/min. in Period 3 PE class
3 | 4 5 6 8
4 | 0 1 1 2 2 3 3 5 5 6 6 7 7 8 9
5 | 0 0 1 6 7 8
6 | 4
    
```

3 | 4 represents 34 sit-ups

Interpreting the data:

The key descriptive terms, such as symmetry, skewness, mode, outliers, clusters and gaps, can all be used with a stem-and-leaf plot, just as with a line plot or histogram. For this example, scores are relatively symmetric with no unique mode. But since most of

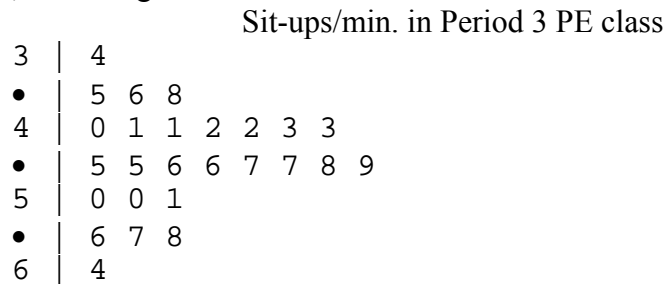
the scores are in the 40s, we could call this the modal class or category. Where is Eric's score? Right in the middle of this modal category. Eric is "typical" of a student from this class.

Variations and extensions:

For children who have mastered the concept of place value it is possible to introduce variants on the stem-and-leaf plot.

a) Split stems:

In the above example, there were many data points in the middle categories. Spreading the data out might reveal additional patterns. One way to do this is to split the stems into a smaller number of equally-sized units. For example, units of 5 ones (30-34, 35-39, 40-44, etc.). Taking the previous plot, we then get:



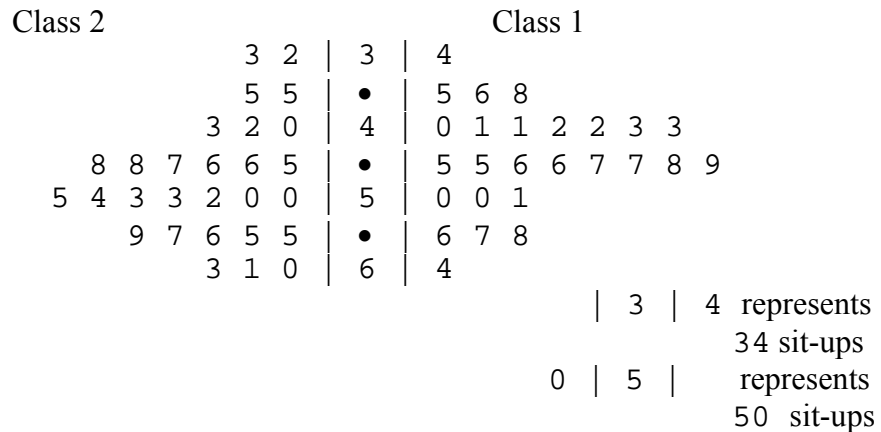
3 | 4 = 34 sit-ups
 • | 5 = 35 sit-ups
 (• is same value as stem above it)

Can you see how this spread out the data?

b) Back-to-back stem-and-leaf plots:

Like line plots and histograms, stem-and-leaf plots can be used to compare two situations. For example, there might be data for a second class, perhaps of older children, or from a class which has been doing sit-ups every day for a couple of weeks:

Comparison of sit-ups/min in 2 classes



Which class is generally more fit?

- ⇒ Note that back-to-back stem-and-leaf plots require writing the left-hand numbers backwards, and therefore this method of plotting is not suitable for the younger children.

- ⇒ Note also that an easy way to demonstrate the back-to-back version is to use 3 transparencies: one containing only the stem, the second, overlaid on the first for the data from situation 1, and the third, overlaid on the first (after removing the second) for the data on situation 2. Then you put all 3 transparencies on top of each other, except that you flip the top one over so the numbers are written backwards. This is suitable for an initial demonstration, or for a quick data collection later, but if you want to keep the plot around for further discussion, you should transfer it onto a single sheet with the digits all written in the normal fashion.

- ⇒ Note that for stem-and-leaf plots you retain the actual values of the numbers. The median number of sit-ups in class 1 can be determined to be 45, while the median in class 2 is 50 sit-ups.

Histograms

Description:

A **histogram** is similar to both a line plot and a stem-and-leaf plot. A line plot shows the distribution of the data over a portion of the real number line—the non-negative integers—where we place a dot or other mark above the appropriate numbers. All such dots above one point on the number line represent exactly one integer. A stem-and-leaf plot looks very similar, except that the “pile” of marks above each point on the number line now represents a range of numbers. In the previous stem-and-leaf plot example, the individual marks were the final digits of the numbers being plotted. If we simply replace these digits with identical marks, and replace the stem indicators with the midpoint of the ranges, we change the stem-and-leaf plot to a histogram. A histogram is not restricted to integer data, however. When there are integer numbers of sampled units (e.g., children in a class) measured for a continuous trait such as height where the data gathered may not be representable as integers (e.g., $35\frac{3}{4}$ ”, 89.5cm), a histogram may be a convenient way to represent the data.

Histograms should not be used with the youngest children (K-1). When children are comfortable counting by 2s, 5s, or other groupings, then histograms can, and should, be introduced. Histograms should be introduced after line plots.

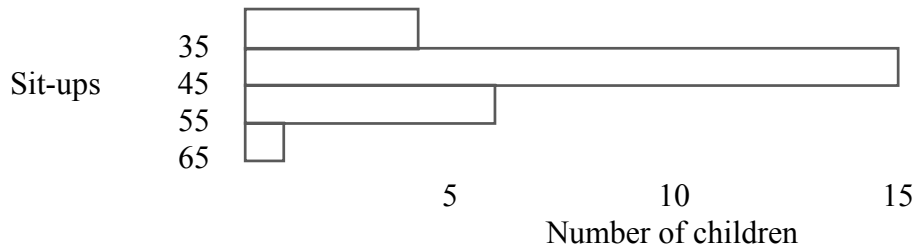
An example:

Consider the original sit-up example for the period 3 PE class from the stem-and-leaf plot.

Sit-ups/min. in Period 3 PE class

Sit-ups	Number of children
35	X X X X
45	X X X X X X X X X X X X X X
55	X X X X X X
65	X

One can also replace the individual marks with a bar which surrounds each “pile” of marks:



⇒ Note that the area in each bar is proportional to the number of data points in each bar because the intervals breaking up the numbers of sit-ups are of equal size.

⇒ Note that the data points need not be integers because the data are grouped into intervals. For example, we could plot heights of children in a class. Height is usually measured in fractional inches (e.g., $35\frac{3}{4}$ "") or in decimal notation (89.5cm).

⇒ Note that there is no "magic" rule that dictates the size of each interval used to group the data. Too many intervals will spread out the data points too much, obscuring a clean picture of the shape of the distribution of the data. Too few will put almost all the points in one interval, also obscuring the shape of the distribution of the data. Students may discover this if they compare graphs that they make from the same data, and if you guide the discussion appropriately.

⇒ Note that the graph can be constructed either horizontally or vertically.

The method of construction:

The steps in constructing a histogram are as follows:

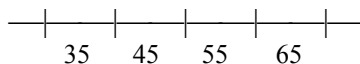
Step #1 Determine the minimum and maximum measured variable. Divide this interval into a number of segments of equal length. It may be convenient to expand the minimum and maximum to slightly more convenient numbers (for example, to the nearest whole or half number).

Step #2 Choose a scale for the Y-axis (including how many counts = 1 dot/post-it if you are using these to plot your results).

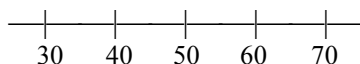
Step #3 Either
a) construct a bar for each interval that is proportional to the data count within each interval, or
b) place one dot/post-it per group of data counts above each appropriate interval.

Initial Preparation:

The preparation for a histogram is essentially the same as for a line plot. Get a piece of butcher paper and place a strip of masking tape horizontally or vertically down one edge of the paper. Mark off the tape in equal-sized intervals. If you are going to use post-its to construct the graph, the intervals should be the width of one post-it. If the interval sizes have already been determined, you can label the intervals, either by writing the midpoints of the interval centered next to the appropriate interval:



or by marking the end points of the intervals:



Variations and extensions:

Two-Way Tables

Description:

When two categorical variables are to be compared, a natural way to summarize the frequency counts is in a **two-way table**, with the categories for one variable forming the column headings, and the categories for the other forming the row headings. For example, students might be categorized by whether they are in elementary or middle school, and whether they are a member of a sports team, or alternatively, which type of sports team are they on.

An example:

Is there a difference in the sports participation of students in middle vs. elementary school? We could create a table which summarizes these two variables for a group of students from one neighborhood as shown below.

		school level		
		elementary	middle	total
sports team	yes	5	15	20
membership	no	5	4	9
	total	10	19	29

Interpreting the Data:

At first glance it may appear that students who do vs. don't participate in sports are equally represented in the two school levels. But we cannot directly compare these numbers because more students are involved in sports than not. We can compare the proportions, however. The proportion of sports participants who are in elementary school is $5/20 = 0.25$, while the proportion of non-participants is $5/9 = 0.56$. The fact that these proportions differ suggests that non-participants are more likely to be in elementary school than are participants. Another way to say this is that the two categories may be *associated* with each other. The proportions calculated for a particular row or column of a table are called the conditional (or marginal) proportions.

Variations and extensions:

These same ideas of conditional proportions and association can be applied to larger and more complex tables. For example, we could break the team membership category into a few sub-categories:

		school level		
		elementary	middle	total
sports	basketball	0	6	6
team	baseball	2	6	8
	soccer	2	3	5
	none	5	4	9

Now we see that 0/6 basketball players are in elementary school, suggesting that this sport may not be accessible to the elementary school students. Note, however, that for this example, breaking up the sports categories leaves us with fairly small members in each cell. As with many other ways of summarizing data, it is desirable to avoid splitting categories too finely—the goal is to maintain at least modest numbers of observations in most categories.

Initial Preparation:

Preparation for collecting data in a 2-way table requires a piece of butcher paper on which you draw a large grid with the appropriate number of rows and columns to reflect the number of categories needed for two types of categorical data. Label the columns and rows. If you will be collecting data points in the form of, e.g., post-its, small photos, or other objects, make sure the squares in the grid are large enough to collect a number of such points.

Graphing:

If students are relatively new to this kind of data display, you may want to have them post dots or post-its in the appropriate squares first. Then, when all students have posted a marker, count the markers and replace with a single number in each cell. Add the numbers in the cells for each row and write the sums to the right of each row. Repeat with the columns, writing the numbers below each column. Place the sum of all cells in the lower right hand corner, to the right of the column totals, and below the row totals.

Scatterplots and correlation

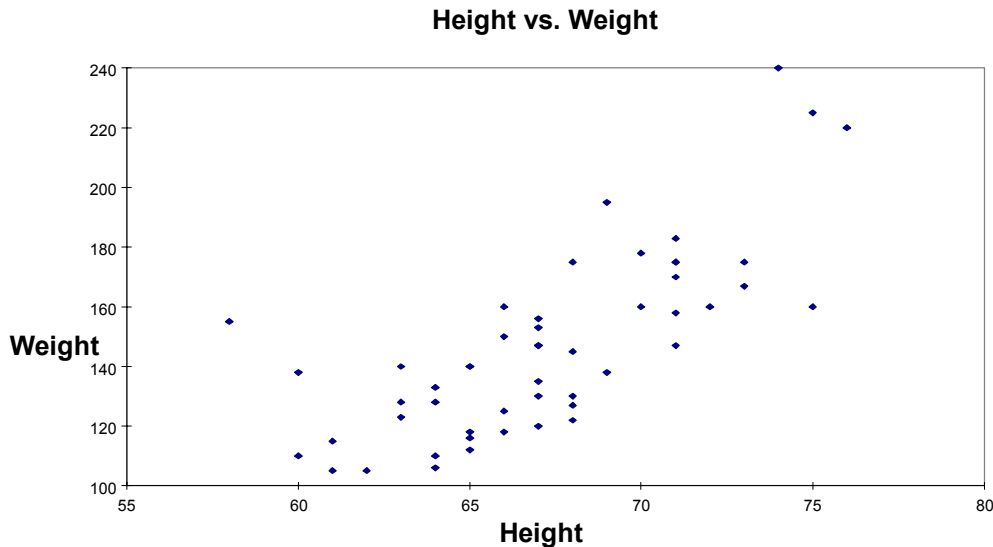
Description:

Scatterplots can be thought of as an extension of a two-way table for measurement of (numeric) data that comes in the form of ordered pairs. An ordered pair consists of a pair of numbers for one individual or object, such that the two numbers are two different measurements on the same individual. Two examples might be height and weight of an individual, or circumference and volume of a sphere. The data could be entered into a very large two-way table with one unit of measurement for each category, but this would not give us a very informative picture of the data. Instead, we can plot a point for each pair of data. The composite data set then can be displayed as a scatterplot.

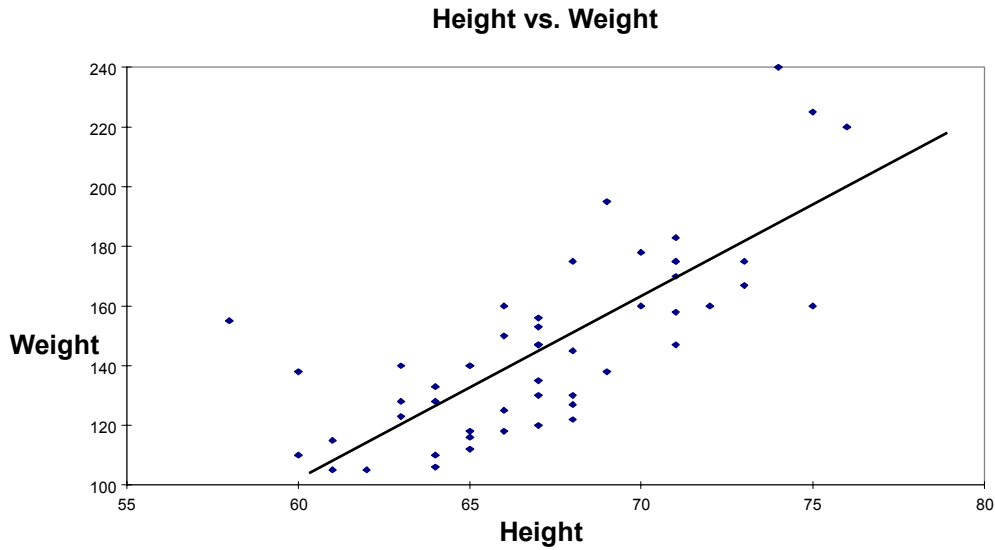
Scatterplots are difficult for the younger children (K-1 or K-2) to use for actual data analysis. However, with appropriate activities (e.g., plotting number of whole peanuts vs. peanuts inside the shell in a handful), scatterplots can be introduced to the younger children. An earlier introduction will increase familiarity with interpretation of the plot so that by about second or third grade it should be possible to begin to use scatterplots for some routine data analyses.

An example:

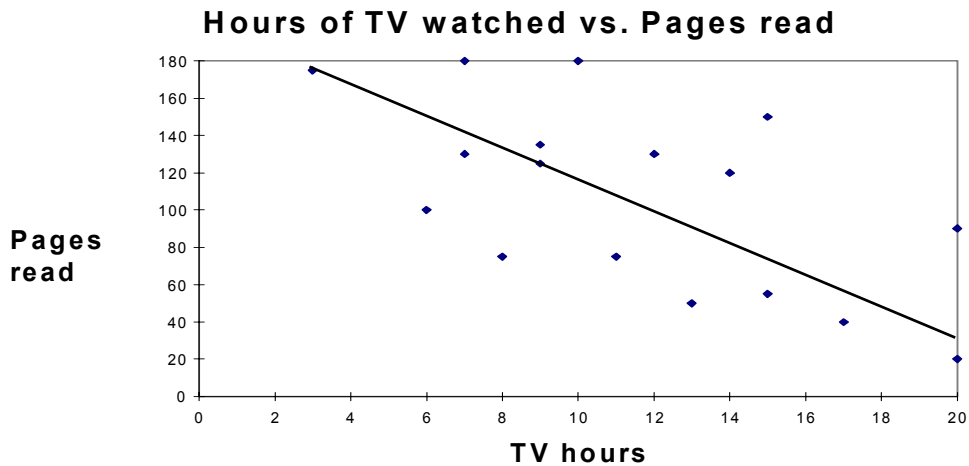
Suppose we have measured the height (in inches) and weight (in pounds) of students in a class. A scatterplot of these data is shown below. Notice that the plot slants upwards as we move from lower to higher heights. This is an example of a positive correlation.



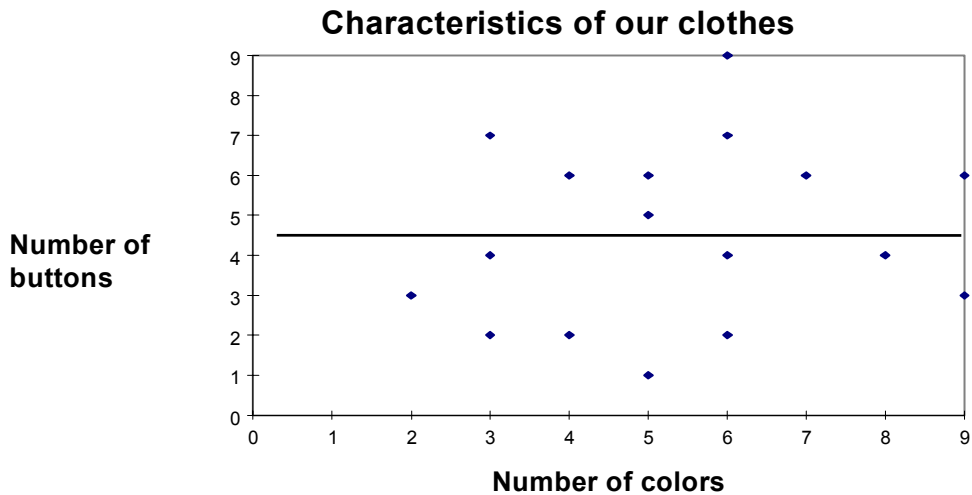
We could further emphasize this association by drawing a line that is as close as possible to all the points; this line would have a positive slope (in more advanced classes formal procedures for choosing this line are introduced, but for elementary and middle-school children a line “by eye” is adequate). We might also emphasize the shape by drawing an ellipse around the majority of the points. This ellipse would also slant upwards.



A second scatterplot might show number of hours of TV watched in a week vs. number of book pages read. This might show a negative association, although with probably more scatter than seen for the height/weight example. If we drew a line or an ellipse around these points, it would slant downwards. Note that we would leave a few outliers outside the ellipse.



Finally, a third scatterplot might show no positive or negative association; in this case we would say there is little or no correlation between the two variables. An ellipse drawn around these points would not slant either upwards or downwards.



Interpreting the data:

As in all other cases, a good way to think about data interpretation is to look for patterns and departures from patterns. In the example of height vs. weight, the overall pattern is that weight increases with height. The line drawn through the points summarizes this increasing trend in much the same way a mean or a median captures the center of a distribution plotted via, e.g., a line plot.

The particular departures from the pattern may also be interesting to consider. Is there something truly unusual about these points, or are they just errors in plotting the data? Two examples of unusual points are the single individual in the center of the weight distribution with very low height, and the 3 points clustered to the far right and above of the rest of the points. Are these data-measurement errors or are they really extreme points (outliers).

Initial preparation:

To make a scatterplot with data collected by students, start with either large sheets of quad-ruled chart paper, or just tape a set of axes on the board or a piece of butcher paper. Find the minimum and maximum values in the class for each variable. Mark off the axes in even intervals to reflect these values, starting at or slightly below the minimum value. For very young children stick with integers, but for older children the numbers can include non-integer scales as well. Label the axes.

Graphing:

Have students come up in small groups (3-5) at a time and place a dot at the intersection representing their pair of numbers. It is useful to model the process with one child first if the children have not done this kind of plot before. To get appropriate placement of dots, have the children first find the correct values for their horizontal point. Then have them place their fingers on this point and trace upwards until they find the vertical-axis value corresponding to their data. That is where they must place their dots.

Version: July 2003

If you want to bring out additional features of the data, for example, the difference between boys and girls in their height/weight relationship, you can give the students different colored dots.

Box Plots

Description:

Box plots provide a visual representation of a five-number summary of data, consisting of the median (the midpoint of the data range), the upper and lower quartiles (the numbers below the highest quarter of the data and above the lowest quarter, respectively) and the largest and smallest values (the extremes). Box plots are particularly useful for comparing distributions of the results from several experimental conditions.

Because box plots are based on simple summaries, they can be used with fairly young children—certainly third graders, but even, in some cases, younger children. For the youngest children, one can “lead up” to a box plot by using only the median (middle number) and the whiskers to the extremes.

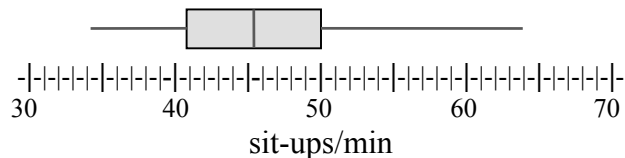
Box plots are an important type of graph to use with children. More than any other type of graph, they focus the user on several key statistical concepts, perhaps the most important of which is that the data can be summarized. Because box plots focus on representing summaries of the data, the children are not distracted by issues such as gaps or multiple modes. In addition, the box plot is a superb way of emphasizing and representing the variability inherent in real data in a way that is computationally accessible to children. This is a key concept—knowing that there is a way of describing or summarizing variability is very important, and will lay the conceptual groundwork for other summaries that can be introduced in high school. Finally, the box plot strongly emphasizes the idea of the center of the distribution. Again, this is a key concept, especially as children begin to compare results from different experimental situations.

An example:

Return to the original sit-ups/min. data set from the stem-and-leaf plot section. The five-number summary, written in the order of the lower extreme, lower quartile, median, upper quartile, and upper extreme is:

34 41 45.5 50 64

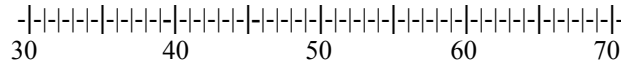
A box plot of these numbers, plotted above the real number line for reference, would look like the following:



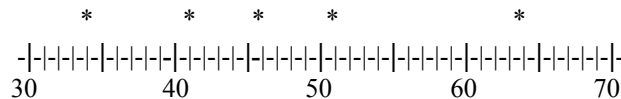
Constructing the graph:

Step #1 Determine the five-number summary.

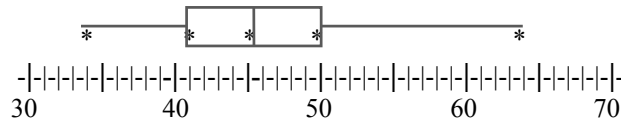
Step #2 Construct a number line that includes the extremes.



Step #3 Mark the position of the five numbers in the summary a little above the number line.

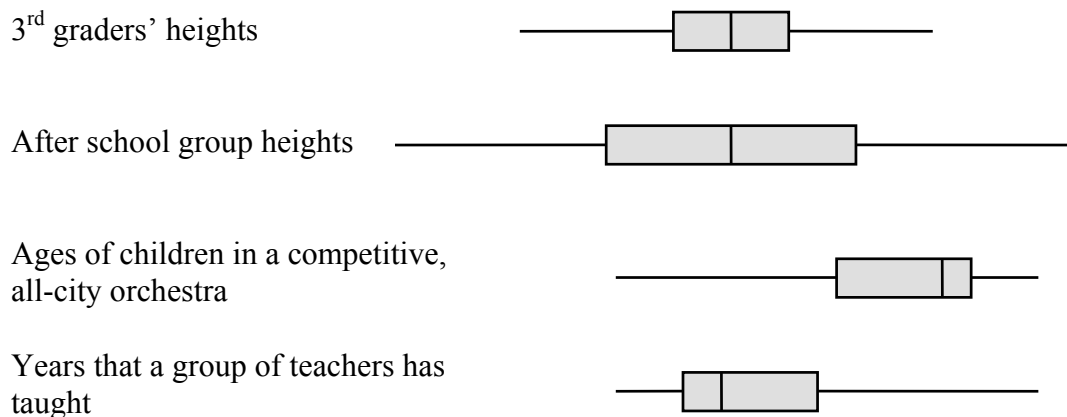


Step #4 Draw a narrow box that connects the quartiles. Draw a line through the box at the median. Draw lines from the ends of the boxes to the extremes (“whiskers”).



Interpreting the data:

Half the scores will lie inside the box. Half the scores will lie outside the box on the whiskers. Half the scores lie below the line in the box, and half lie above this line. One quarter of the scores lie on each whisker and in each half of the box. Below are some additional examples, without the number lines (and plotted on different scales), to illustrate possible shapes of resulting distributions:

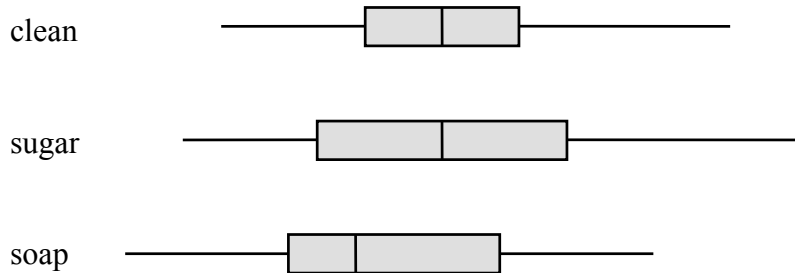


Variations and extensions:

Parallel box plots provide a convenient way to compare different situations. It is easy to compare more than two conditions, and it is easy to determine where salient features of the data lie. Parallel box plots are also a convenient way to compare two (or more)

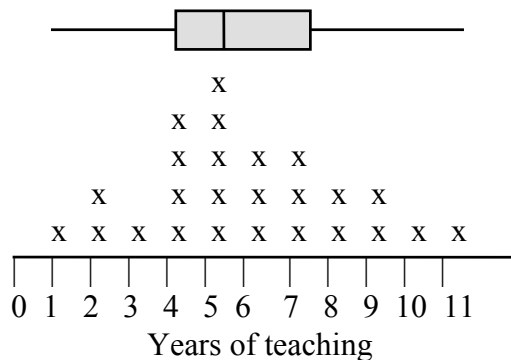
groups with very different sample sizes (in this situation, back-to-back line plots or histograms are hard to compare).

For example, children have counted the number of drops of clean, soapy, and sugary water that fit on pennies:



One can see that there is not much of a difference in the centering of the clean vs. sugary water, although the clean water example may be a little less variable. One can also see the downward shift in the number of drops of soapy water, although there is still a lot of overlap with the other two conditions.

When making box plots, it is often useful to make them after first plotting the data as a line plot, or a stem-and-leaf plot. Then you plot the box plot next to the same set of axes:



This is particularly useful when you first introduce the box plot as it will relate a more familiar type of graph to this new type of graph. If you are comparing 2 (or more) conditions, you will find that the box plot probably will bring out the location of the distribution better than will the link plot; by putting the box plots next to the line plots, this point will come out quite clearly, especially if there is a slight shift in the distribution.

Time-Series Plots

Description:

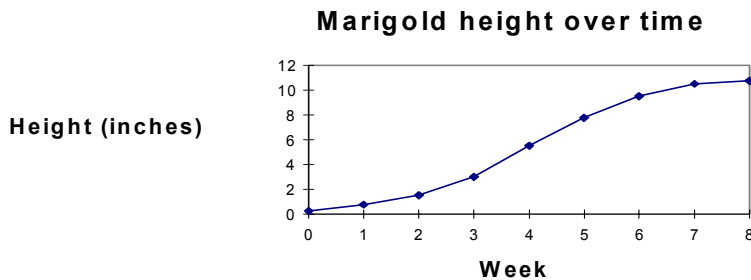
Where a single individual or object has been measured for some characteristic at multiple time points, a **time-series plot** is often used to show the results. Some examples where this type of plot may be informative are: height of a child at different times over several years, air temperature each morning over a month, or number of seeds sprouted each day in a model ecosystem over a week. Time-series plots can be used in the youngest grades, as long as the units of measurement fall into a range with which the children are comfortable.

An example:

Mary measured the height, in inches, of a marigold plant for 8 weeks from the time the plant first came up. Her data:

week:	0	1	2	3	4	5	6	7	8
height (in):	0.25	0.75	1.5	3.0	5.5	7.75	9.5	10.5	10.75

When she plotted the results, they looked like the following graph:



Initial Preparation:

If you have measured one object as a class over time then you probably will make a single, large graph using quad-ruled graph paper. Alternatively, if individuals or teams of children have each measured an object or individual over time, then they will need to make individual graphs, unless you want them to all plot their results on one large class graph. This should only be done if there is only a limited number of such plots to be placed on the graph, e.g., about 6 or fewer.

Determine the smallest and largest measurement collected over the time of the study. These will determine the scale of the dependent axis (the y-axis or vertical axis). In the data above, the measured heights run from 0.25 inches to 10.75 inches; a reasonable scale for these data would be $\frac{1}{2}$ inch for each square of $\frac{1}{4}$ " graph paper (this will allow the graph to fit on one page, while providing sufficient separation vertically to see the increase in height over time). On the dependent axis, we have time of 0 to 8 weeks. We could use 4 squares/week for the time axis. Mark off the time units on the horizontal axis

(for time-series plots, time is almost always put on the horizontal axis). Mark off the measurement units on the vertical axis. Label the two axes and add a title.

Graphing:

Plot the data, using dots, the same way as one would for scatterplots: find the point on the horizontal axis first, and use a finger to trace upwards until the appropriate point on the vertical axis is reached. Place a dot at this intersection. Connect the dots with straight lines.

Variations and extensions:

You can compare multiple plots by putting them in the same graph, using different colors or line-styles to distinguish them. If you do this, remember to add a key. For example, one might grow one marigold under each of three different conditions, and plot the growth of each plant on the same graph.

It can be quite instructive as an exercise to plot the same data on two graphs that have the same scale on one axis (e.g., the horizontal axis), but very different scales on the other axis. Try doubling or halving the size of the units. Clearly, the two graphs represent the same data, but one's initial interpretation may be very different. If individuals in a class are allowed to choose their own scales to plot data from a class data set this point will probably emerge naturally as you ask the children to compare their plots. But you may need to guide the discussion of this issue, and to suggest or request that some children replot their data on a scale which is the same as that used by other children to make the point.

Comparisons among Graph types

What is the distinction between bar plots, line plots, stem-and-leaf plots, and histograms? When is it appropriate to use one but not another?

While each of these four graphing styles can be used to summarize integer counts of data, each has its own advantages and disadvantages in representing different kinds of information. Bar plots, commonly recommended in exercises aimed at introducing the concept of graphs to kids, are really only appropriate for summarizing data sorted by a categorical (non-numerical) trait, such as color (“blue” does not logically come before “green” and after “red”). For data which are naturally classified by number, however, it would be more useful to use one of the other three graphing styles. For instance, a line plot or a stem-and leaf plot is appropriate if your data set is relatively small (~25-35) and consists of integer values. If your data is grouped by intervals of numbers (i.e., 0-9, 10-19, 20-29 etc.) or consists of non-integers (e.g., $4\frac{3}{4}$, 2.5), you should use a histogram or a stem-and-leaf. A histogram is convenient if your data set is large, because you can pool by larger intervals, and you do not need to plot individual points. A histogram may also be useful when the range of data is very high, but the number of points is modest; by grouping data into intervals, some picture of the shape of the distribution may emerge.

When should one use a box-plot instead of a histogram, line plot, or stem-and-leaf plot?

When you want to emphasize summary statistics associated with a data set (e.g., the median and some measure of variability) then the box plot will do a better job at conveying this information than will one of the other plots. In addition, when you want to compare several data sets, parallel box plots are often a better choice than back-to-back or separate histograms. In the context of science experiments, where children are trying to make inferences about the effects of certain experimental conditions, parallel box plots are more likely to provide a clear picture of a change in distribution than are histograms. However, box plots do not allow you to see such aspects of the distribution as where there are gaps or clusters, whether there are multiple peaks in the distribution, or other such details. Box plots are also better for large rather than small data sets (e.g., more than 20 or so data points).

When should you connect the dots on a plot (e.g., when should you use a time-series plot instead of a scatterplot)?

Generally, you should only connect the dots when the horizontal axis represents time, *and* the points at the different times were measured on the same subject. The line makes the implicit assumption that there was a measurement between the two points being collected. If the data plotted at different time points are from different individuals, then you should just represent the results as a scatterplot *without* connecting the dots. For example, if one plots the diameter of a single evaporating circle of water over time (minutes), then one could use a time-series plot. If Mary plots her results at 1 minute, Joseph plots his at 2 minutes, and so on throughout the class, a scatterplot would be a better choice.

When should you use a scatterplot to compare two variables as opposed to a back-to-back line plot (or histogram or stem-and-leaf plot)?

If you have collected two numerical conditions on each individual (e.g., test scores before and after a lesson on a topic), the scatterplot allows you to determine whether there is an association (positive or negative) between the two scores. A back-to-back plot loses this information. With a back-to-back plot we cannot tell if a person with a high score under one condition also gets a high score under the second condition. However, if you want to see if two experimental conditions produce different distributions—either in the center of the distribution or the variability—back-to-back or parallel box plots may bring out the effect more clearly than a scatterplot. It *is* possible to make this determination from a scatterplot by noting whether or not most of the points are above (or below) the 45 degree line, but it is harder to see the effect than with the aforementioned styles.