1. We use graphs to communicate data– pictographs, pie charts, bar graphs, line graphs....
The PICTOGRAPH is perhaps the most fun since it offers a creative opportunity in the choice of symbol.
For example, in charting the popularity of songs I chose the symbol ♪....

Notice the ← TITLE and LEGEND

FAVORITE SONGS OF 40 KINDERGARTNERS

If You're Happy (and)	♪ ♪ ♪ ♪ ♪ ♪ ♪ ♪
Three Blind Mice	♪ ♪ ♪
Old MacDonald	♪ ♪ ♪ ♪
The Hokey Pokey	♪ ♪ ♪ ♪ ♪

(Each ♪ represents 2 votes)

2. Picto-graph the cookie drive: Mr. Jones' class sold 150 boxes; Ms. Smith's, 180; M. Durite's, 220.

⇨

A pictograph is appropriate for displaying data when:

⇨

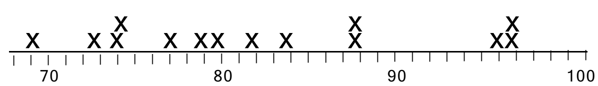
On the NSAT (National Science Achievement Test),
Ms. Smith's science class made the following scores:

69	73	74	74	77	79	80
82	84	88	88	96	97	97

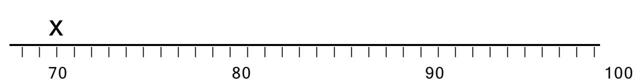
& Mr. Jones' class earned these:

85	73	70	98	83	75	76
97	82	72	83	72	84	84

3. Here is a LINE PLOT of Ms. Smith's class scores



⇨ Make a line plot of Mr. Jones' class scores



⇦

4. Ms. Smith's class scores in a STEM-AND-LEAF diagram are shown below.

Classify all the data in a back-to-back stem-and-leaf diagram below.

Scores of Ms. Smith's 8th-graders on the National Science Achievement Test

⇨
⇨
⇨

6	9
7	3 4 4
7	7 9
8	0 2 4
8	8 8
9	
9	6 7 7

Legend:
7 | 7 9
represents
scores of
77 & 79

Note: did not use JUST FOUR classes such as 70-79, 80-89.... Four is considered too few classes!

Scores of two 8th-grade classes on National Science Achievement Test

Smith's Class		
9	6	
4 4 3	7	
9 7	7	
4 2 0	8	
8 8	8	
	9	
7 7 6	9	

⇦

Legend: ⇦

______
______
______
______
______

Ms. Smith's science class made the following scores: & Mr. Jones' class earned these:

69	73	74	74	77	79	80	85	73	70	98	83	75	76
82	84	88	88	96	97	97	97	82	72	83	72	84	84

5. A FREQUENCY TABLE lists ranges of values for the data, and their frequencies- the number of data that fall in each range . Classify the data in a combined frequency table. (Use classes that correspond to the stem-and-leaf diagram above.)

⇨ Title?

Scores on test	Number of students
65-69	1
70-74	7
75-79	4
80-84	

⇨

6. Show the combined data in a HISTOGRAM using classes which correspond to those in problem (3). A histogram uses adjacent rectangles on a Cartesian coordinate system to display data. The horizontal or x-axis displays the values of the data; the vertical or y-axis the frequencies.

USE the SPACE!!

⇨
⇨
⇨

Histograms and Bar Graphs show frequencies (vertical axis*) of data grouped in categories (horizontal axis), in summary form. The DIFFERENCE is that the HISTOGRAM is used when the range of possible values for data is CONTINUOUS, whereas a BAR GRAPH is used when the possible data values are SEPARATE VALUES, rather than a continuous range. For example:
We show distributions of trees BY HEIGHT via a histogram, since tree heights cover a continuous range.
We show distribution of trees BY TYPE (oak, sycamore, manzanita) on a bar graph.
⇨ Show HOW MANY BOXES OF COOKIES SOLD BY EACH CLASSROOM at Elm St. Elementary on a _____________.

(Bar graphs are often displayed sideways, with the variable of interest on the vertical axis and frequencies on the horizontal axis. Histograms are generally not drawn sideways.)

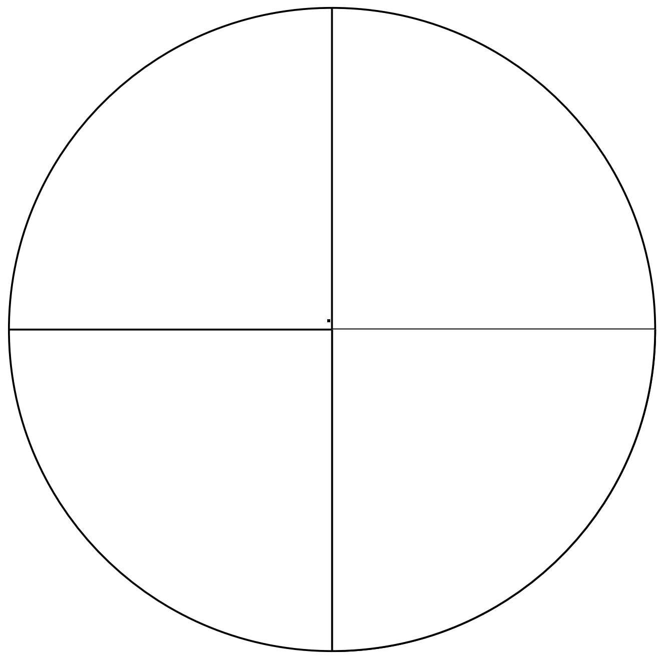
The areas of the rectangles (or bars) must be in proportion to the frequencies with which data falls into each category. In histograms, generally the classes, or categories, comprise equal ranges of possible data values, except when there is a good reason to do otherwise.

7. Gina spent the following amounts every month on the average while attending United University in ‘96.
Illustrate the proportions with a PIE CHART (CIRCLE GRAPH).
Don’t forget titles and legends.
Label each sector/segment ($ amt or %)

⇨

ITEM	$$$	degrees
Rent	$300	$^{300}/_{900} \times 360°$
Food	100	
Books	50	
Tuition	400	
Clothing & misc	50	
Total		

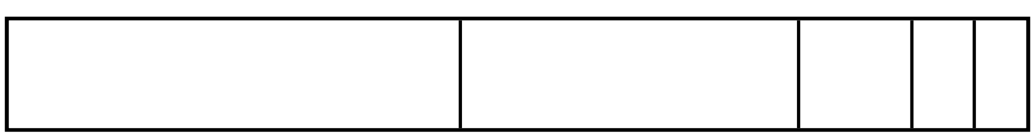
⇨
⇨



A “single bar graph” does the same thing as a pie chart, but in a bar rather than a “pie”. For the same data given above, complete this:

⇨

⇨



"A STATISTIC" can be a value calculated from data that represents some characteristic of the data. Among statistics commonly used to describe the "average" (or "typical") value of a set of numeric data are the MEAN, the MEDIAN and the MODE.

8. The MEDIAN of data is the value at which 50% of the data consists of higher values, and 50% lower.
   - In short: The dividing line between the top half and the bottom half. The value in the middle. (The average of the two values in the middle, when the number of data is even.)

   Using an ordered stem-and-leaf diagram, for instance, we can easily ascertain the median.

⇨ The median of Ms. Smith's science class scores is:

⇨ The median of Mr. Jones' science class scores is:

⇨ The median of the combined data is:

9. The MODE is the most frequent value in the data. Find the mode of the combined science test scores.

⇨

If there are two equally most-frequently-occurring values, we say the distribution is bimodal, and has two modes. If there are three, then we have to call it trimodal. (Four? Let's not go there!)

10. The MEAN is the arithmetic average: MEAN = (sum of data)/(number of data)
    Compute the mean of Ms. Smith's class, using the raw data from #3.

⇨

11. Suppose the person in Ms. Smith's class who scored 69 had instead given up and received a score of 0. What would the class mean have been?

⇨

12. If M. Durite's 28 students achieved a mean of 70, what is the combined mean for the two classes?
    (It is NOT 76.35 !)

⇨

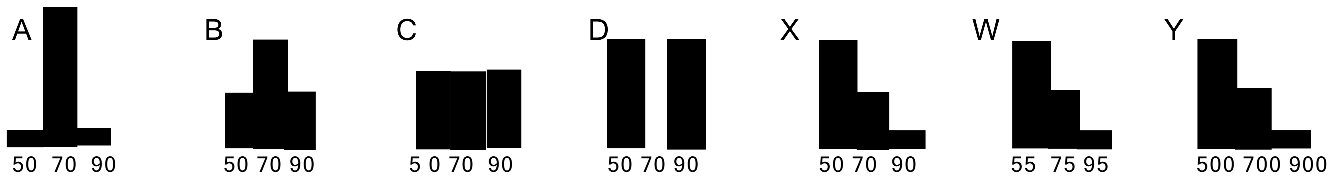
13. Suppose two students in your class earn 100 on a test, five earn 85, and thirteen earn 70.

⇨ Would the mean score for those students on that test be (100+85+70)/3 = 255/3 = 85?

14. Following are mini-histograms for very simplistic sets of grouped data– kept very simple so that we can gain some understanding of the mean and the median.

For each distribution, find the median & the mean:

a	f	b	f	c	f	d	f	x	f	w	f	y	f
50	1	50	3	50	4	50	6	50	6	55	6	500	6
70	8	70	6	70	4	70	0	70	3	75	3	700	3
90	1	90	3	90	4	90	6	90	1	95	1	900	1



⇨
mean: $\bar{a} = 70$ $\bar{b} =$ $\bar{c} =$ $\bar{d} =$ $\bar{x} =$ $\bar{w} =$ $\bar{y} =$

median: $\tilde{a} =$ $\tilde{b} =$ $\tilde{c} =$ $\tilde{d} =$ $\tilde{x} =$ $\tilde{w} =$ $\tilde{y} =$
⇨

15. The first four distributions are symmetric.
⇨ When the distribution is symmetric, the mean & median are ...

The fifth (x) distribution is asymmetric, "skewed to the right" (because the “tail” is on the right).
⇨ Where is the mean, relative to the median?

Note: When data is skewed, it is because some values are exceptionally high (skewed right) or low (skewed left). A few exceptionally high values have a profound effect on the mean, but no effect on the median, as we will see on the next page. Thus data skewed right will have a mean higher than median, because the mean is pulled high by the assymetrically high values- high values that are not balanced by corresponding low values.

16. The values in the W distribution are each 5 more than those in the X distribution ($w_1 = 5 + x_1$, et cetera).

⇨ How do the means for W & X compare? [Rhetorical question: Does this make sense?]

IF EVERY VALUE IN A SET OF DATA IS INCREASED BY AN AMOUNT "A" (ie. AMOUNT A IS ADDED), THEN THE MEAN IS
⇨

17. The values in the Y distribution are each 10 times those in the X distribution ($y_1 = 10x_1$, $y_2 = 10x_2$, etc).

⇨ How does the mean for Y compare with the mean for X ? [Does this make sense?]

IF EVERY VALUE IN A SAMPLE OR POPULATION IS MULTIPLIED BY A FACTOR "R", THEN THE MEAN IS
⇨

18. Suppose the example X data above is doubled, then increased by 3. What is the new mean?
⇨

At right are the selling prices of the 30 single-family homes sold in Northridge in January, 1995 (1 yr. after...).

$	f
125K	15
175K	11
250K	2
1000K	1
3000K	1

19. Find the median.
⇨

20. Is the mean higher or lower than the median?
Find it.
⇨

21. Find the mode. (OK, the “modal class”)
⇨

22. If you are interested in the price of housing in a particular area, which of the above "average" statistics would you want to know, to estimate the price of houses in that area? Why?
⇨

Statistics that tell us the “typical” value of a set of data:

Mode = most frequent value

Median = the middle score [the (n+1)/2th] $\tilde{x}$

Mean = the evenly distributed total; also the balancing point of the distribution $\bar{x}$ or μ

What’s the difference between $\bar{x}$ and μ ?

Statisticians use $\bar{x}$ to refer to the mean of a sample (taken from a population),
and μ for the mean of the entire population .

Measures of Central Tendency: Among statistics commonly used to describe the amount of "variation", or spread, in a set of data... are the STANDARD DEVIATION, the RANGE and the INTERQUARTILE RANGE.

The RANGE is the total span or spread of the data, i.e. the highest value - the lowest value.

Just as the median divides the ordered data into two equal groups, in order, QUARTILE MARKS divide the data into four equal groups. These quartile marks are referred to as the first and third quartile marks, and the second quartile mark, which is also the median.

23. Compute the range and interquartile range of the scores in Mr. Jones' class.

➩ range = maximum data value − minimum data value = 28

➩ Interquartile range = $Q_3 - Q_1$ = 11

70 72 72 73 75 76 82 | 83 83 84 84 85 97 98

The standard deviation is a bit more complicated....

STANDARD DEVIATION = sq. root of (average square distance from the mean) = $\sqrt{\dfrac{\Sigma(x-\mu)^2}{n}}$ ← μ *or* $\bar{x}$; ← n, *or* n − 1 for a sample

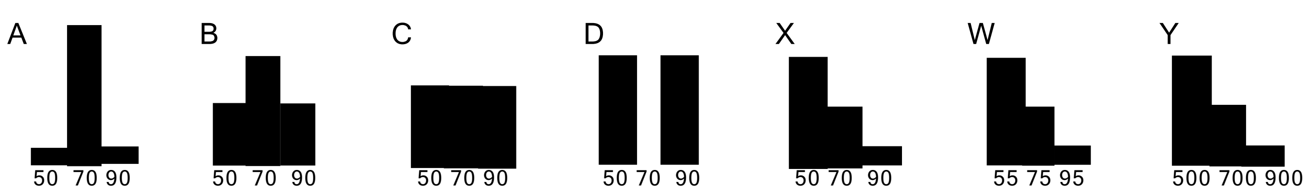
24. Compute the s.d. (s or σ) for the data: 1 2 3 4 5 (First find the mean!)

x	x - mean distance	(distance)^2
1		
2		
3		
4		
5		
	Total:	

➩ Would the data: 1 1 3 5 5 have the same standard deviation as the data above?

25. Consider the simple grouped data examples we used to explore the mean and median— we compute the standard deviations:

a	f	b	f	c	f	d	f	x	f	w	f	y	f
50	1	50	3	50	4	50	6	50	6	55	6	500	6
70	8	70	6	70	4	70	0	70	3	75	3	700	3
90	1	90	3	90	4	90	6	90	1	95	1	900	1



mean:	$\bar{a}$ = 70	$\bar{b}$ = 70	$\bar{c}$ = 70	$\bar{d}$ = 70	$\bar{x}$ = 60	$\bar{w}$ = 65	$\bar{y}$ = 600
➩	sd ≈ 8.96	sd ≈	sd ≈ 16.3	sd ≈ 20	sd ≈ 13.4	sd ≈ 13.4	sd ≈ 134
	s ≈ 9.4		s ≈ 17.1	s ≈ 20	s ≈ 14.1	s ≈ 14.1	s ≈ 141

26. Would you say the first four distributions are "clustered about the mean" to the same degree, or are some more "spread out"? How is this reflected by the standard deviation?
⇨
⇨

27. What kind of data would have a standard deviation 0 ? ...a negative standard deviation ?
⇨
⇨

28. The values in the W distribution are each 5 more than those in the X distribution, as noted earlier.
⇨ How does the standard deviation for W compare with the std. deviation for X? Does this make sense?
⇨

29. The values in the Y distribution are 10 times those in the X distribution, as noted earlier.
⇨ How does the standard deviation for Y compare with the std. deviation for X? Does this make sense?
⇨

30. If every value in a sample or population is MULTIPLIED by a factor "r", then
⇨ the mean is

⇨ the standard deviation is

If an amount "a" is ADDED to every value in a sample or population, then
⇨ the mean is

⇨ the standard deviation is

Statistics that tell us about the amount of "spread" in the data:

Range = highest value – the lowest value = the "width" of the data

Interquartile range = third quartile – first quartile = the "width of the middle 50%" of the data

Standard deviation = square root of average square distance from the mean (almost)