1. About Data

- *VARIABLE*
- *VALUE*
- *TYPES OF VARIABLES*
  - Quantitative
    - Discrete
    - Continuous
  - Qualitative
- *THE DISTRIBUTION (PATTERN) OF DATA*

2. The Histogram -- a way to examine distributions

Examples are on pages 30 & 31 of your text.

A HISTOGRAM is a graph that shows percentages by area. The rectangles are called "bins." Bins are first constructed by creating intervals. The key to a histogram is that it is the AREA of the "bin", **not** the height of the "bin", that is important. The area of the bin is proportional to the relative frequency of observations in the bin: The area of each "bin" is proportional to the number of observations in the class interval. The total area MUST be 100%. Specifically the area in each bin is:

(#observations in the bin)/(total number of observations) and remember area=base*height

The histogram has class intervals on its horizontal axis. This axis must be scaled. The histogram does not require a scale on its vertical axis. The vertical axis is scaled in percent per unit of the horizontal axis, and a scale is automatically imposed by the fact that the area of the histogram **must** be 100%

We also need an "endpoint convention" to be able to draw a histogram: if an observation falls on the boundary between two class intervals, to which one should we associate it? The two standard choices are always to include the left boundary and exclude the right, except for the rightmost bin, or always to include the right boundary and exclude the left, except for the leftmost bin.

3. Building a Histogram

a. First sort the data in increasing order
b. Choose class intervals
c. Count the number of observations that fall into each class interval
d. Construct rectangles with observations proportional to the relative frequencies (percentages) with which the observations fall in each interval

There are no standard rules for determining appropriate class intervals, and the impression one gets of how the data are distributed depends on the number and location of the intervals.
Histograms are usually used for large datasets.

4. Things to be aware of with respect to histograms

- *Center*
- *Symmetry or skewness*
- *Spread*
- *Exceptions ("outliers")*
- *Remark:* the histogram can take all kinds of shapes.

5. Bad Graphics

The whole point of this exercise is really to help you learn how to convey information in a meaningful way using graphics.

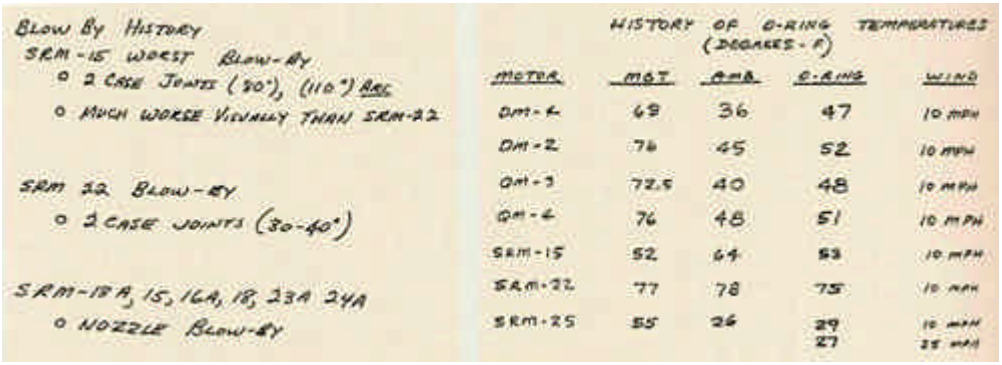
http://www.usatoday.com/snapshot/money/msnap119.htm
http://www.usatoday.com/snapshot/news/nsnap194.htm
http://www.usatoday.com/snapshot/news/snapndex.htm
http://lilt.ilstu.edu/gmklass/pos138/datadisplay/badchart.htm

6. Why this matters

Good graphics allows us to see patterns, trends, or other structures that would be hidden if drawn poorly. The story behind the space shuttle Challenger is perhaps the saddest chapter of statistical graphics.

The Space Shuttle Challenger exploded shortly after take-off in January 1986. Subsequent investigation determined that the cause was failure of the O-ring seals used to isolate the fuel supply from burning gases.

This is data in it's rawest form. An engineer's notes with some organization.



Blow By History
SRM-15 worst Blow-By
- 2 case joints (50°), (110°) Rec
- much worse visually than SRM-22

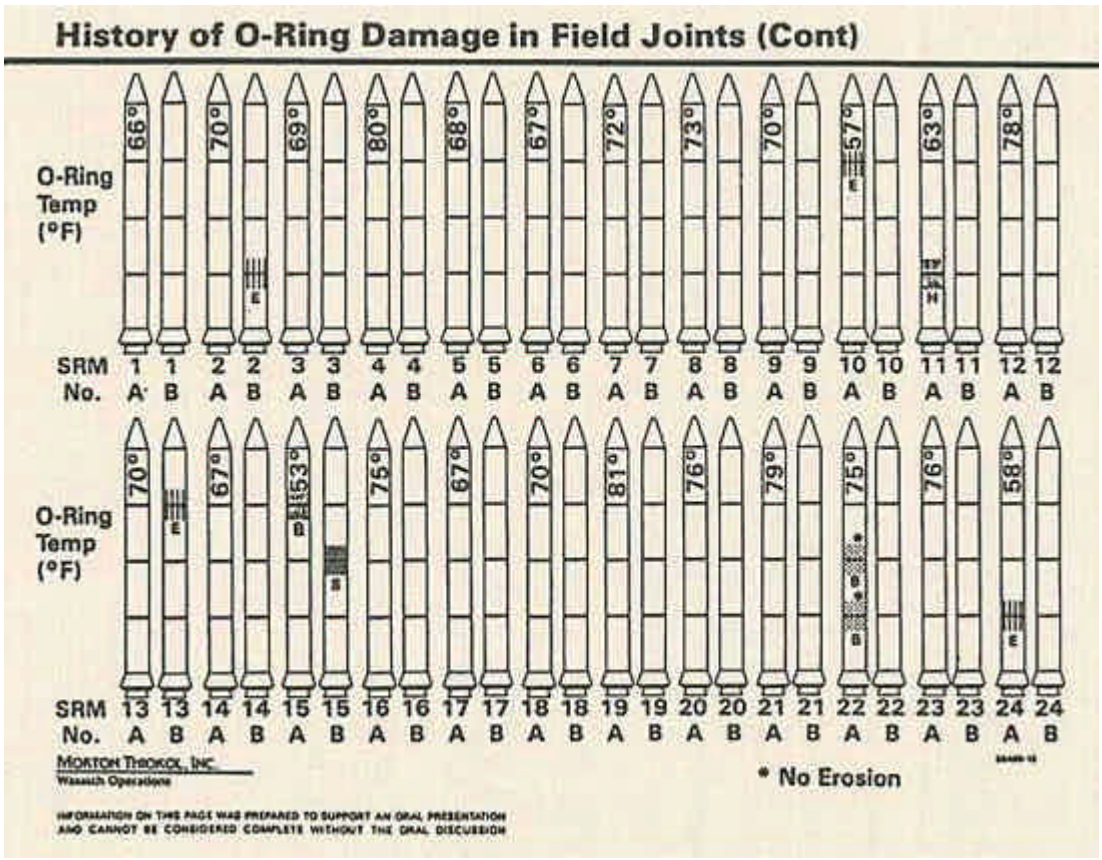
SRM 22 Blow-By
- 2 case joints (30-40°)

SRM-18A, 15, 16A, 18, 23A 24A
- nozzle Blow-By

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

MOTOR	MBT	AMB.	O-RING	WIND
DM-1	69	36	47	10 MPH
DM-2	76	45	52	10 MPH
DM-3	72.5	40	48	10 MPH
DM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

This is the data as summarized by the engineers at Morton Thiokol and presented to their supervisors.



The same information, reorganized. The temperature on the launch pad that morning was about 29 degrees Fahrenheit.