21. Assuming that "start the first class at 200 lb" refers to the first lower class limit produces the frequency table below.

weight (lbs)	frequency
200 - 219	6
220 - 239	5
240 - 259	12
260 - 279	36
280 - 299	87
300 - 319	28
320 - 339	0
340 - 359	0
360 - 379	0
380 - 399	0
400 - 419	0
420 - 439	0
440 - 459	0
460 - 479	0
480 - 499	0
500 - 519	1
	175

In general, an outlier can add several rows to a frequency table. Even though most of the added rows have frequency zero, the table tends to suggest that these are possible valid values -- thus distorting the reader's mental image of the distribution.

23. The two frequency tables are given below.

a.

height	frequency
66 - 67	4
68 - 69	3
70 - 71	10
72 - 73	10
74 - 75	0
76 - 77	1
	28

b.

height	frequency
66 - 67	6
68 - 69	4
70 - 71	4
72 - 73	4
74 - 75	4
76 - 77	6
	28

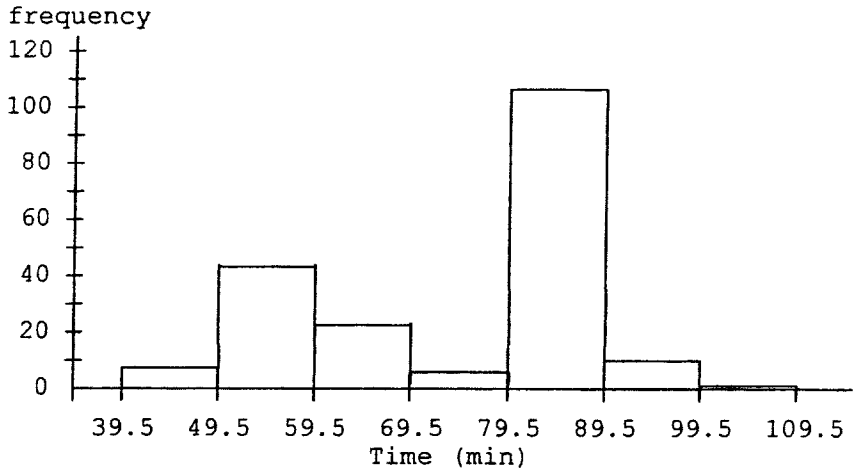
Data set (b) appears to be the phony data for two reasons. (1) The frequencies in set (b) follow a regular pattern unlikely to be achieved by chance, while the frequencies in set (a) follow the type of irregular pattern expected by chance. (2) The pattern in (b) [heights fairly uniformly distributed with more at the extremes than near the middle] disagrees with the generally accepted pattern in (a) [many heights near the middle values and fewer at the extremes].

2-3 Pictures of Data

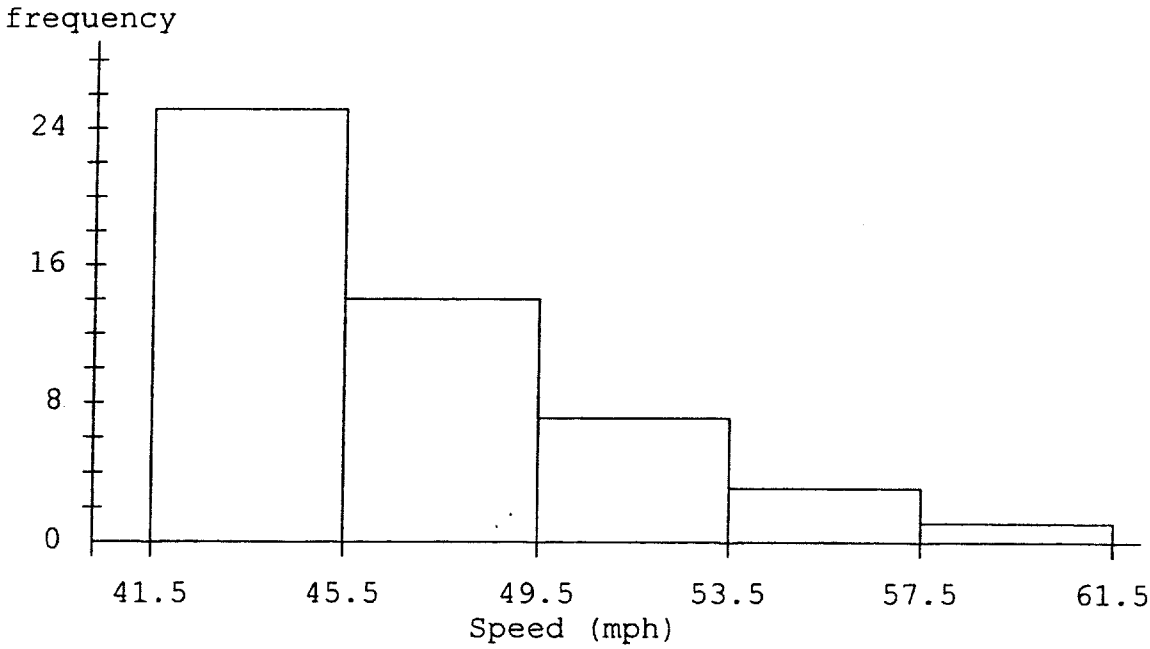
1. 42, the height of the bar centered at 0.0

3. 2, the one Monday represented by the bar centered at 1.0 and the one Monday represented by the bar centered at 1.4

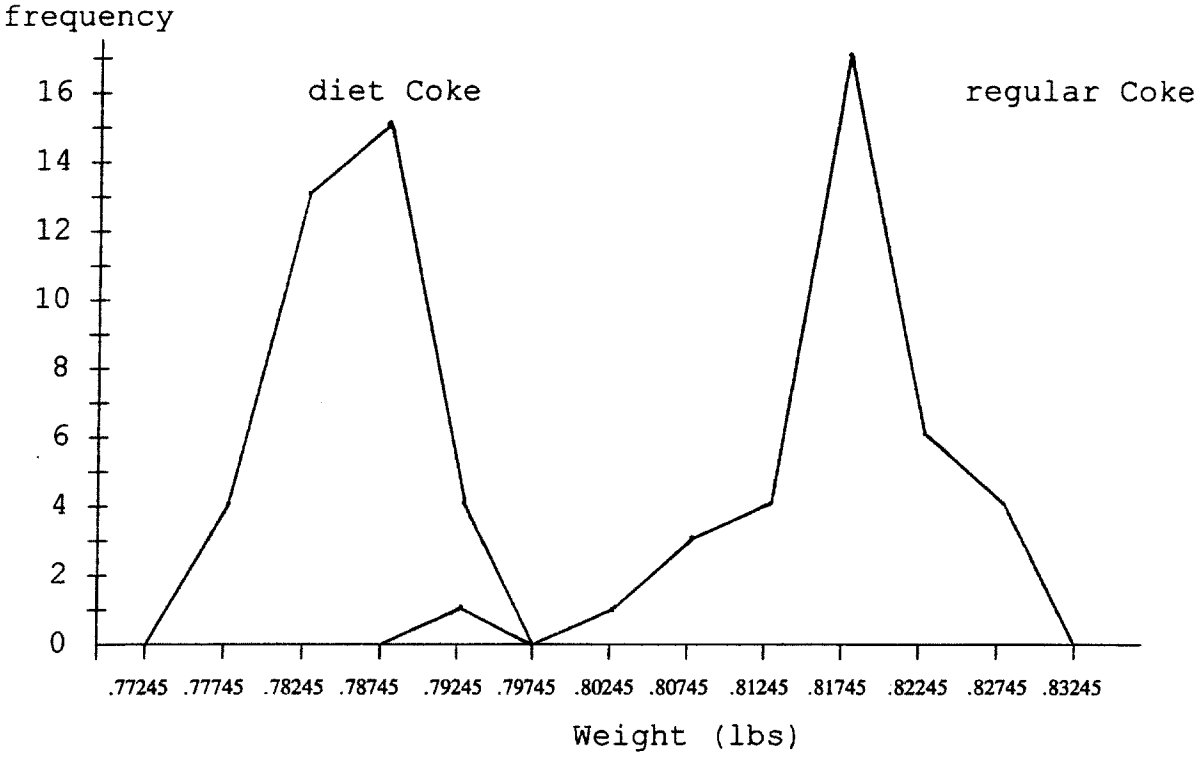
5. See the figure below. The bars extend from class boundary to class boundary. Each axis is labeled numerically and with the name of the quantity represented. Barring an interval longer than any previously recorded, a minimum stay of 109.5 minutes assures seeing an eruption. A minimum stay of 99.5 minutes includes 199 of the 200 (i.e. 99.5%) recorded intervals and would be inadequate only if longest interval occurred *and* the tour arrived within the first ten minutes of that interval -- or about (1/200) x (10/110) = 0.00045 = .045% of the time, about once in every 2200 tours.



7. See the figure below. The bars extend from class boundary to class boundary. Each axis is labeled numerically and with the name of the quantity represented. Although the posted limit is 30 mph, it appears that the police ticket only those traveling at least 42 mph.

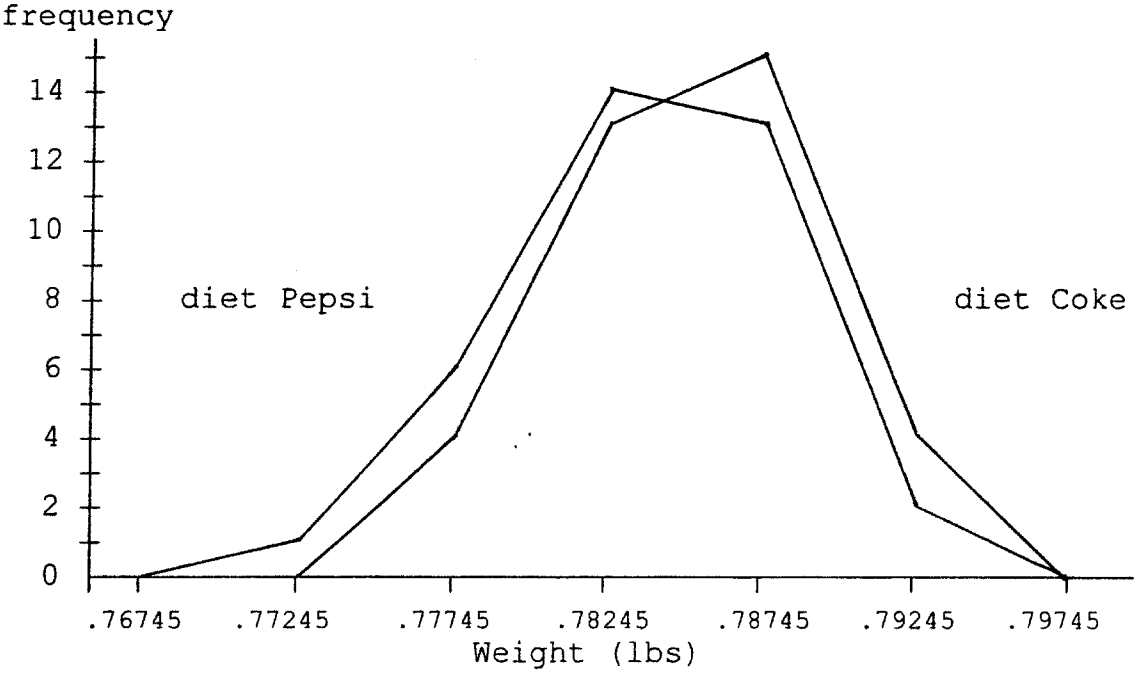


9. See the figure below. The frequencies are plotted above the class midpoints, and "extra" midpoints are added so that both polygons begin and end with a frequency of zero.



The two distributions have similar shapes, but the diet Coke weights are less than those of regular Coke.

11. See the figure below. The frequencies are plotted above the class midpoints, and "extra" midpoints are added so that both polygons begin and end with a frequency of zero.

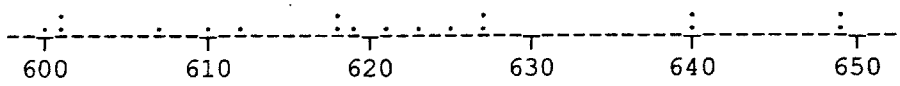


The two distributions are very similar, but the diet Pepsi weights are slightly smaller than those of diet Coke.

13. The original numbers are listed by the row in which they appear in the stem-and-leaf plot.

stem	leaves	original numbers
60	0117	600, 601, 601, 607
61	02889	610, 612, 618, 618, 619
62	13577	621, 623, 625, 627, 627
63		
64	0099	640, 640, 649, 649

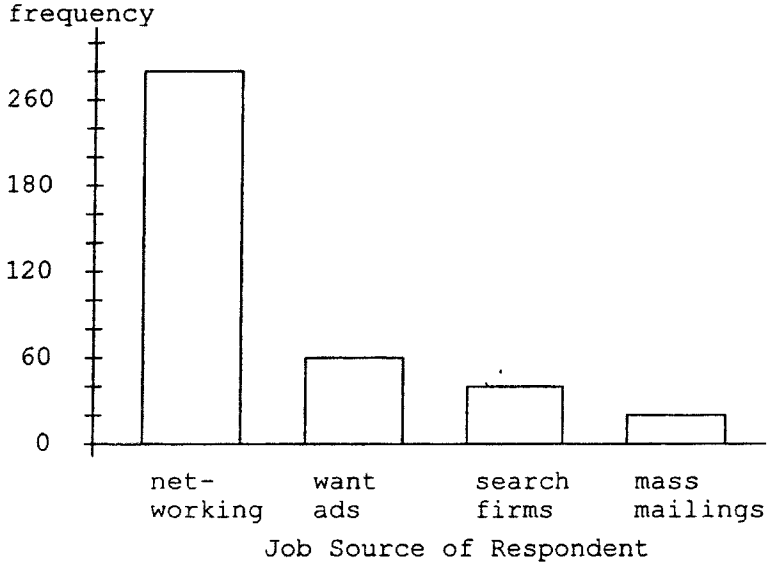
15. The dotplot is constructed using the original scores as follows.



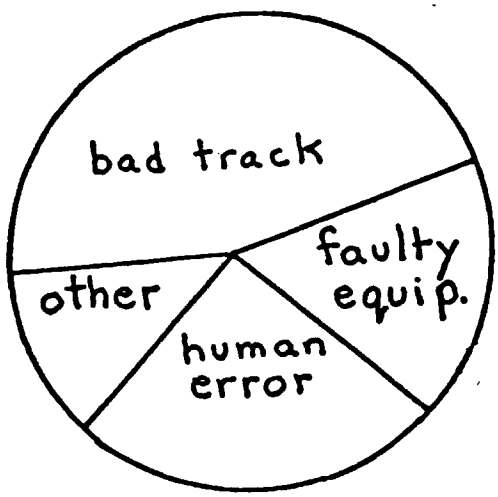
17. The expanded stem-and-leaf plot below on the left is one possibility. NOTE: The text claims that stem-and-leaf plots enable us to "see the distribution of data and yet keep all the information in the original list." Following the suggestion to round the nearest inch not only loses information but also uses subjectivity to round values exactly half way between. Since always rounding such values "up" creates a slight bias, many texts suggest rounding toward the even digit -- so that 33.5 becomes 34, but 36.5 becomes 36. The technique below of using superscripts to indicate the occasional decimals is both mathematically clear and visually uncluttered.

stem	leaves
3	$6\ 7$
4	$0\ 0\ 1\ 3\ 3^5$
4	$6\ 6\ 7\ 8\ 8\ 9$
5	$0\ 2\ 2^5\ 3\ 3\ 4$
5	$7^3\ 7^5\ 8\ 9\ 9\ 9$
6	$0\ 0^5\ 1\ 1\ 1^5\ 2\ 3\ 3\ 3\ 3^5\ 4\ 4\ 4$
6	$5\ 5\ 6^5\ 7\ 7^5\ 8^5$
7	$0\ 0^5\ 2\ 2\ 2\ 2\ 3\ 3^5$
7	$5\ 6^5$

19. See the figure below, with bars arranged in order of magnitude. Networking appears to be the most effective job-seeking approach.

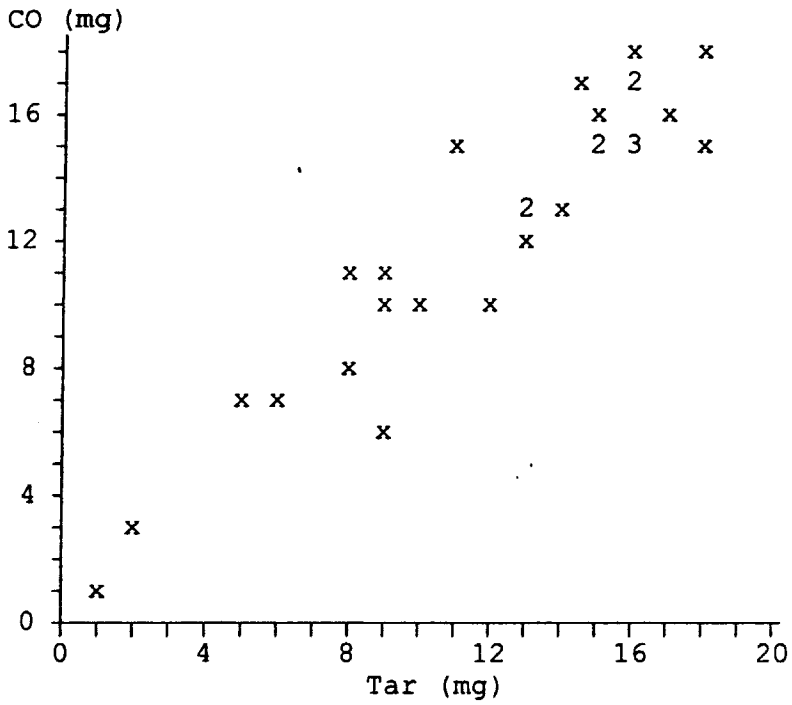


21. See the figure below. The sum of the frequencies is 50; the relative frequencies are 23/50 = 46%, 9/50 = 18%, 12/50 = 24%, and 6/50 = 12%. The corresponding central angles are (.46)360° = 165.6°, (.18)360° = 64.8°, (.24)360° = 86.4°, and (.12)360° = 43.2°. NOTE: To be complete, the figure needs to be titled with the name of the quantity being measured.



Causes of Train Derailments

23. The scatter diagram is given below. The figure should have a title, and each axis should be labeled both numerically and with the name of the variable. An "x" marks a single occurrence, while numbers indicate multiple occurrences at a point. Cigarettes high in tar also tend to be high in CO. The points cluster about a straight line from (0,0) to (18,18), indicating that the mg of CO tends to be about equal to the mg of tar.



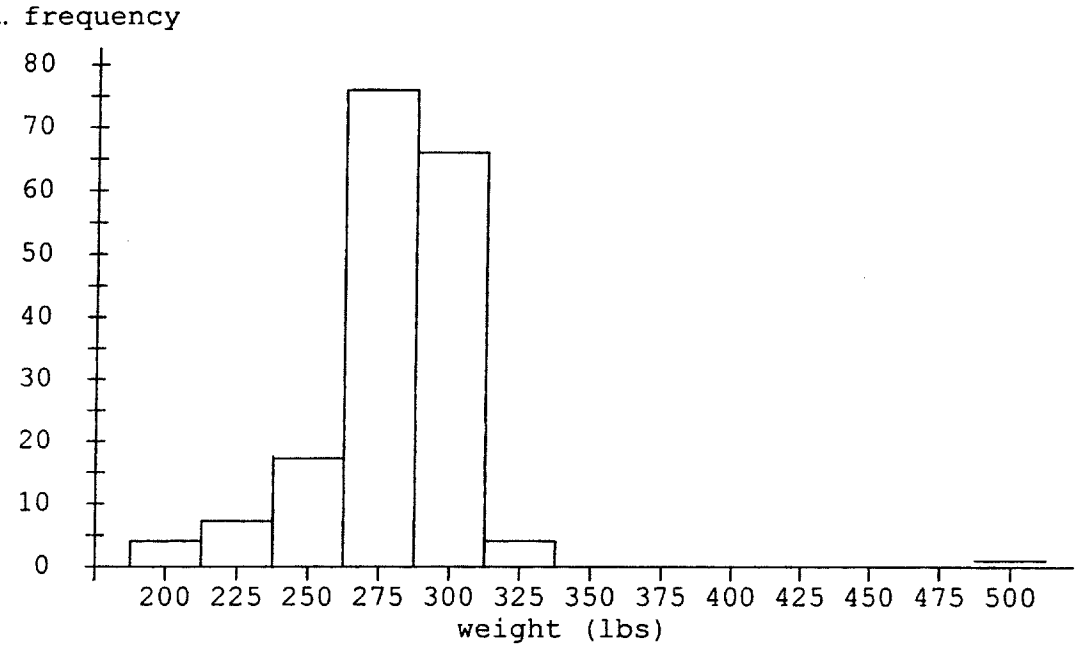
TAR AND CO CONTENT OF SELECTED 100mm FILTERED AMERICAN CIGARETTES

25. According to the figure, 422,000 started and 10,000 returned. 10,000/422,000 = 2.37%

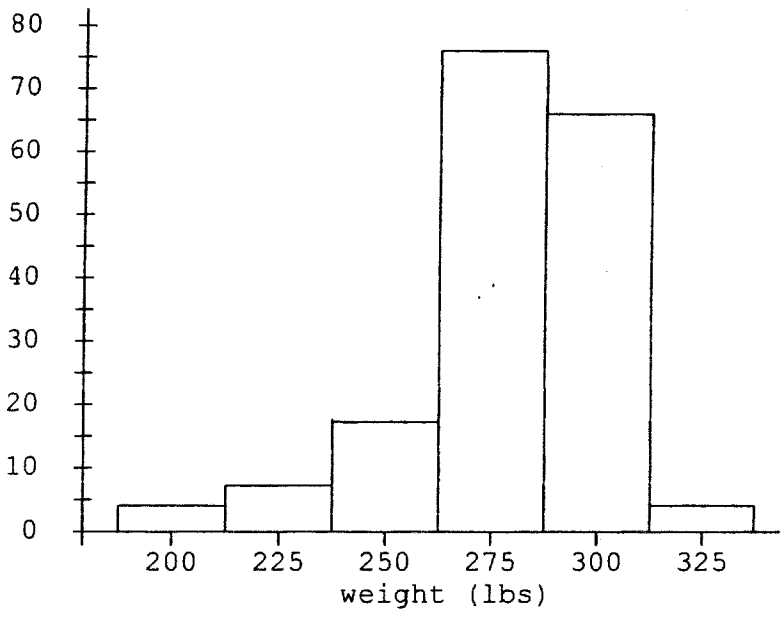
27. The figure indicates the number of men had just dropped to 37,000 on November 9 when the temperature was 16°F (-9°C), and had just dropped to 24,000 on November 14 when the temperature was -6°F (-21°C). The number who died during that time, therefore, was 37,000 - 24,000 = 13,000.

29. NOTE: Exercises #20 an #21 of section 2-2 dealt with frequency table representations of this data and specified a class width of 20. Using a class width with an odd number of units of measure allows the class midpoint to have the same number of decimal places as the original data and often produces more appealing visual representations. Here we use a class width of 25 -- with class midpoints of 200, 225, etc. The histogram bars extend from class boundary to class boundary (i.e., from 187.5 to 212.5 for the first class), but for convenience the labels have been placed at the class midpoints.

a.



b.



c. The basic shape of the histogram does not change, except that a distant piece has been "broken off." NOTE: The redrawn histogram in part (b) should not be an exact copy of the one in part (a) with the distant bar erased. Since removing the distant bar reduces the effective width of the figure significantly, the rule that the height should be approximately 3/4 of the width requires either making the remaining bars wider (and keeping the figure's height) or making them shorter (and keeping the figure's reduced width). To do otherwise produces a figure too tall for its width -- and one that tends to visually overstate the differences between classes.

31. a. The final form of the back-to-back stem-and-leaf plot is given below, an example of adapting a standard visual form in order to better communicate the data. While such decisions are arbitrary, we choose to display "outward" from the central stem but to keep the actors' ages in increasing order from left to right.

actor's age		actress' age
	2	1466678
122235677899	3	00113344445557789
00122334556788	4	111249
13566	5	0
0012	6	011
6	7	4
	8	0

b. Female Oscar winners tend to be younger than male Oscar winners. If one assumes that acting ability doesn't peak differently for females and males, the data may reveal a difference in the standards by which females and males are judged.

2-4 Measures of Center

NOTE: As it is common in mathematics and statistics to use symbols instead of words to represent quantities that are used often and/or that may appear in equations, this manual employs symbols for the measures of center as follows:

mean = $\bar{x}$ mode = M

median = $\tilde{x}$ midrange = m.r.

Also, this manual follows the author's guideline of presenting means, medians and ranges accurate to one more decimal place than found in the original data. The mode, the only measure which must be one of the original pieces of data, is presented with the same accuracy as the original data.

1. Arranged in order, the 12 scores are: 8 11 13 14 14 14 15 16 17 18 25 27

a. $\bar{x} = (\Sigma x)/n = (192)/12 = 16.0$

b. $\tilde{x} = (14 + 15)/2 = 14.5$

c. M = 14

d. m.r. = (8 + 27)/2 = 17.5

NOTE: The median is the middle score when the scores are arranged in order, and the midrange is halfway between the first and last score when the scores are arranged in order. It is therefore usually helpful to begin by placing the scores in order. This will not affect the mean, and it may also aid in identifying the mode. In addition, no measure of central tendency can have a value lower than the smallest score or higher than the largest score -- remembering this helps to protect against gross errors, which most commonly occur when calculating the mean.