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. $\bar{x} = (\Sigma x)/n = (192)/12 = 16.0$ c. M = 14

b. $\tilde{x} = (14 + 15)/2 = 14.5$ d. m.r. = (8 + 27)/2 = 17.5

NOTE: The median is the middle score <u>when the scores are arranged in order</u>, and the midrange is halfway between the first and last score <u>when the scores are arranged in order</u>. 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.

3. Arranged in order, the 12 scores are: 35 46 55 65 74 83 88 93 99 107 108 119
   a. $\bar{x} = (\Sigma x)/n = (972)/12 = 81.0$
   b. $\tilde{x} = (83 + 88)/2 = 85.5$
   c. M = [none]
   d. m.r. = (35 + 119)/2 = 77.0

   Yes; these results are acceptable and consistent with the goal of service in 90 seconds or less.

5. Arranged in order, the scores are as follows.
   JV: 6.5 6.6 6.7 6.8 7.1 7.3 7.4 7.7 7.7 7.7
   Pr: 4.2 5.4 5.8 6.2 6.7 7.7 7.7 8.5 9.3 10.0

<u>Jefferson Valley</u>	<u>Providence</u>
n = 10	n = 10
$\bar{x} = (\Sigma x)/n = (71.5)/10 = 7.15$	$\bar{x} = (\Sigma x)/n = (71.5)/10 = 7.15$
$\tilde{x} = (7.1 + 7.3)/2 = 7.20$	$\tilde{x} = (6.7 + 7.7)/2 = 7.20$
M = 7.7	M = 7.7
m.r. = (6.5 + 7.7)/2 = 7.10	m.r. = (4.2 + 10.0)/2 = 7.10

   Comparing only measures of central tendency, one might suspect the two sets are identical. The Jefferson Valley times, however, are considerably less variable.

   NOTE: This is the reason most banks have gone to the single waiting line. While it doesn't make service faster, it makes service times more equitable by eliminating the "luck of the draw" -- i.e., ending up by pure chance in a fast or slow line and having unusually short or long waits.

7. Arranged in order, the scores are as follows.
   Coke: .8150 .8163 .8181 .8192 .8211 .8247
   Pepsi: .8156 .8170 .8211 .8216 .8258 .8302

<u>Coke</u>	<u>Pepsi</u>
n = 6	n = 6
$\bar{x} = (\Sigma x)/n = (4.9144)/6 = .81907$	$\bar{x} = (\Sigma x)/n = (4.9313)/6 = .82188$
$\tilde{x} = (.8181 + .8192)/2 = .81865$	$\tilde{x} = (.8211 + .8216)/2 = .82135$
M = [none]	M = [none]
m.r. = (.8150 + .8247)/2 = .81985	m.r. = (.8156 + .8302)/2 = .82290

   Pepsi appears to weigh slightly more than Coke. Since the cans are sold by volume and not by weight, this is not a question of ethics -- merely a reflection of the fact that the two drink cans of the same volume have different ingredients.

9. The x values below are the class midpoints from the given frequency table.

x	f	x·f
44.5	8	356.0
54.5	44	2398.0
64.5	23	1483.5
74.5	6	447.0
84.5	107	9041.5
94.5	11	1039.5
104.5	1	104.5
	200	14870.0

$$\bar{x} = (\Sigma x \cdot f)/n = (14870.0)/200 = 74.35$$

   NOTE: The mean time was calculated to be 74.35 minutes. According to the rule given in the text, this value should be rounded to one decimal place. The text describes how many decimal places to present in an answer, but not the actual rounding process. When the figure to be rounded is exactly half-way between two values (i.e., the digit in the position to be discarded is a 5, and there are no further digits because the calculations have "come out even"), there is no universally accepted rounding rule. Some authors say to always round up such a value; others correctly note that always rounding up introduces a consistent bias, and that the value should actually be rounded up half the time and rounded down half the time. And so some authors suggest rounding toward the even value (e.g., .65 becomes .6 and .75 becomes .8), while others simply suggest flipping a coin. In this manual, answers exactly half-way between will be reported without rounding (i.e., stated to one more decimal than usual).

11. The x values below are the class midpoints from the given frequency table.

x	f	x·f
43.5	25	1087.5
47.5	14	665.0
51.5	7	360.5
55.5	3	166.5
59.5	1	59.5
	50	2339.0

$\bar{x} = (\Sigma x \cdot f)/n$
$= (2339.0)/50$
$= 46.8$

The mean speed of 46.8 mi/hr of those ticketed by the police is more than 1.5 times the posted speed limit of 30 mi/hr. NOTE: This indicates nothing about the mean speed of *all* drivers, a figure which may or may not be higher than the posted limit.

13. The following values were obtained for new textbooks, where x_i indicates the ith score from the ordered list.

author's college	UMASS
n = 35	n = 40
$\bar{x} = (\Sigma x)/n = 2016.85/35 = 57.624$	$\bar{x} = (\Sigma x)/n = 2607.70/40 = 65.1175$
$\tilde{x} = x_{18} = 59.35$	$\tilde{x} = (x_{20}+x_{21})/2 = (70.00+72.70)/2 = 71.35$

Assuming a random sample was taken from each institution, textbooks appear to cost more at the University of Massachusetts than at the author's college. This is probably more a reflection of the courses offered at each institution and/or the textbook selection practices of faculty than of bookstore pricing policies.

15. The following values were obtained for Boston rainfall, where x_i indicates the ith score from the ordered list.

Thursday	Sunday
n = 52	n = 52
$\bar{x} = (\Sigma x)/n = 3.57/52 = .069$	$\bar{x} = (\Sigma x)/n = 3.52/52 = .068$
$\tilde{x} = (x_{26}+x_{27})/2 = (.00 + .00)/2 = .000$	$\tilde{x} = (x_{26}+x_{27})/2 = (.00 + .00)/2 = .000$

If "it rains more on weekends" refers to the amount of rain, the data do not support the claim. The amount of rainfall appears to be virtually the same for Thursday and Sunday. If "it rains more on weekends" refers to the frequency of rain (regardless of the amount), then the proportions of days on which there was rain would have to be compared.

17. Let $\bar{x}_h$ stand for the harmonic mean.

$\bar{x}_h = n/[\Sigma(1/x)]$
$= 2/[1/40 + 1/60] = 2/[.0417] = 48.0$

19. R.M.S. $= \sqrt{\Sigma x^2/n}$
$= \sqrt{[(110)^2 + (0)^2 + (-60)^2 + (12)^2]/4} = \sqrt{15844/4} = \sqrt{3961} = 62.9$

21. Let the original data, arranged in order, be: $x_1, x_2, \ldots, x_n$

a. Arranged in order, the new data would be: $x_1+k, x_2+k, \ldots, x_n+k$
In general, adding (or subtracting) a constant k from each score will add (or subtract) k from each measure of center.
new $\bar{x} = [\Sigma(x+k)]/n = [\Sigma x + nk]/n = (\Sigma x)/n + (nk)/n = \bar{x} + k$
new $\tilde{x} = \tilde{x}+k$ [from the ordered list]
new $M = M+k$ [from the ordered list]
new m.r. $= [(x_1+k) + (x_n+k)]/2 = [x_1 + x_n + 2k]/2 = (x_1+x_n)/2 + (2k)/2 = \text{m.r.}+k$

b. Arranged in order, the new data would be: $k{\cdot}x_1, k{\cdot}x_2, \ldots, k{\cdot}x_n$
In general, multiplying (or dividing) each score by a constant k will multiply (or divide) each measure of center by k.
new $\bar{x} = [\Sigma(k{\cdot}x)]/n = [k{\cdot}(\Sigma x)]/n = k{\cdot}[(\Sigma x)/n] = k{\cdot}\bar{x}$
new $\tilde{x} = k{\cdot}\tilde{x}$ [from the ordered list]
new $M = k{\cdot}M$ [from the ordered list]
new m.r. $= (k{\cdot}x_1 + k{\cdot}x_n)/2 = k{\cdot}(x_1 + x_n)/2 = k{\cdot}\text{m.r.}$

NOTE: Do not assume the principle in (a) and (b) extends automatically to other mathematical operations. A short example will show, for example, that the mean of the squares of a set of data is not necessarily the square of their original mean.

23. a. Arranged in order, the original 54 scores are:

26	29	34	40	46	48	60	62	64	65	76	79	80	86	90	94
105	114	116	120	125	132	140	140	144	148	150	150	154	166	166	180
182	202	202	204	204	212	220	220	236	262	270	316	332	344	348	356
360	365	416	436	446	514										

$\bar{x} = (\Sigma x)/n = (9876)/54 = 182.9$

b. Trimming the highest and lowest 10% (or 5.4 = 5 scores), the remaining 44 scores are:

48	60	62	64	65	76	79	80	86	90	94	105	114	116	120	125
132	140	140	144	148	150	150	154	166	166	180	182	202	202	204	204
212	220	220	236	262	270	316	332	344	348	356	360				

$\bar{x} = (\Sigma x)/n = (7524)/44 = 171.0$

c. Trimming the highest and lowest 20% (or 10.8 = 11 scores), the remaining 32 scores are:

79	80	86	90	94	105	114	116	120	125	132	140	140	144	148	150
150	154	166	166	180	182	202	202	204	204	212	220	220	236	262	270

$\bar{x} = (\Sigma x)/n = (5093)/32 = 159.2$

In this case, the mean gets smaller as more scores are trimmed. In general, means can increase, decrease, or stay the same as more scores are trimmed. The mean decreased here because the higher scores were farther from the original mean than were the lower scores.