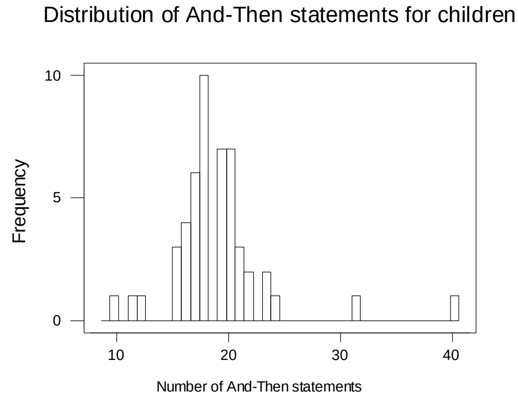
Chapter 2 - Describing and Exploring Data

2.1 Children's recall of stories:

a.

Children's "and then...s"	Frequency
10	1
11	1
12	1
15	3
16	4
17	6
Num18	10
19	7
20	7
21	3
22	2
23	2
24	1
31	1
40	1



b. unimodal and positively skewed

2.3 The problem with making a stem-and-leaf display of the data in Exercise 2.1 is that almost all the values fall on only two leaves if we use the usual 10s' digits for stems.

```
Stem   Leaf
  1    01255566667777778888888888899999999
  2    000000011122334
  3    1
  4    0
```

And things aren't much better even if we double the number of stems.

```
Stem  Leaf
 1*   012
 1.   555666677777788888888889999999
 2*   000000011122334
 2.
 3*   1
 3.
 4*   0
```

Best might be to use the units digits for stems and add HI and LO for extreme values

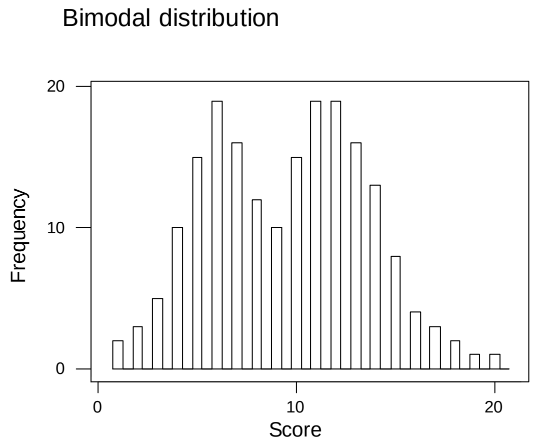
Stem	Leaf
5	555
6	6666
7	7777777
8	8888888888
9	9999999
10	0000000
11	111
12	22
13	33
14	4
HI	31 40

2.5 Stem-and-leaf diagram of the data in Exercises 2.1 and 2.4:

Children		Adults
	0*	1
	0t	34
	0f	55
	0s	7777
	0.	88889999999
10	1*	00000000111111
2	1t	222223
555	1f	4444555
7777776666	1s	667
77777778888888888	1.	
1110000000	2*	
3322	2t	
4	2f	
	2s	
	2.	
40 31	Hi	

2.7 Invented bimodal data:

Score	Freq
1	2
2	3
3	5
4	10
5	15
6	19
7	16
8	12
9	10
10	15
11	19
12	19
13	16
14	13
15	8
16	4
17	3
18	2
19	1
20	1



2.9 The first quartile for males is approximately 77, whereas for females it is about 80. The third quartiles are nearly equal for males and females, with a value of 87.

2.11 The shape of the distribution of number of movies attended per month for the next 200 people you met would be positively skewed with a peak at 0 movies per month and a sharp dropoff to essentially the baseline by about 5 movies per month.

2.13 Stem-and-leaf for ADDSC

Stem	Leaf
2.	69
3*	0344
3.	56679
4*	00023344444
4.	5566677888899999
5*	00000000011223334
5.	55677889
6*	00012234
6.	55556899
7*	0024
7.	568
8*	
8.	55

2.15 **a.** $Y_1 = 9 \;\; Y_{10} = 2$

b. $\Sigma Y = 9 + 9 + \ldots + 2 = 57$

2.17 **a.** $\Sigma Y^2 = 9^2 + 9^2 + \ldots + 2^2 = 377$

b.
$$\frac{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}{N-1} = \frac{377 - \frac{3249}{10}}{9} = 5.789$$

c. $\sqrt{\text{answer to Exercise 17b}} = \sqrt{5.789} = 2.406$ **c.**

d. The units of measurement were squared musicality scores in part (b) and musicality scores in part (c).

2.19 **a.**
$$\Sigma(X+Y) = (10+9) + (8+9) + \ldots + (7+2) = 134$$
$$\Sigma X + \Sigma Y = 77 + 57 = 134$$

b.
$$\Sigma XY = 10(9) + 3(8) + \ldots + 3(7) = 460$$
$$\Sigma X \Sigma Y = (77)(57) = 4389$$

c.
$$\Sigma CX = \Sigma 3X = 3(10) + 3(8) + \ldots + 3(7) = 231$$
$$C\Sigma X = 3(77) = 231$$

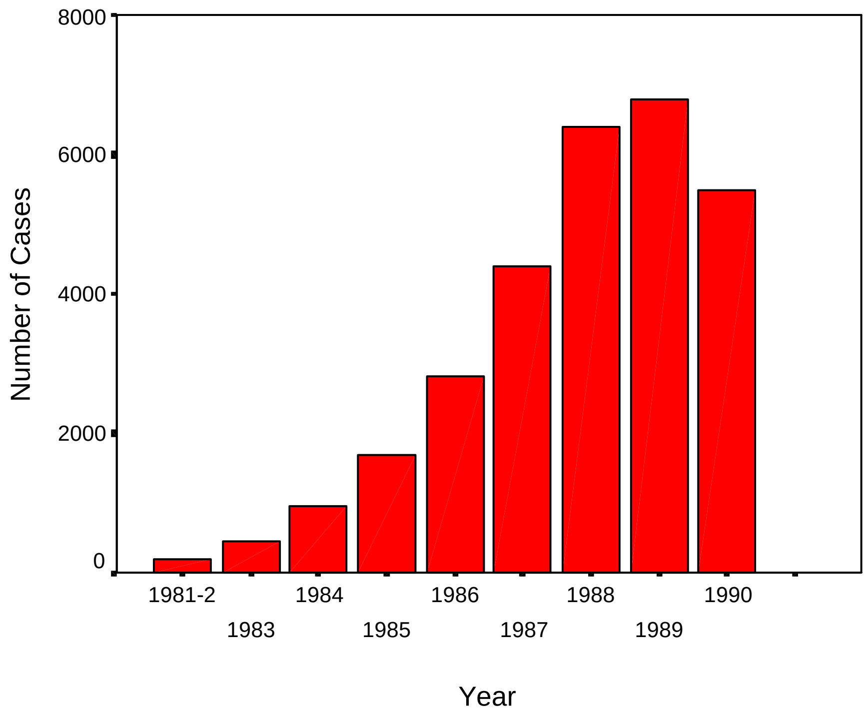
d.
$$\Sigma X^2 = 10^2 + 8^2 + \ldots + 7^2 = 657$$
$$(\Sigma X)^2 = 77^2 = 5929$$

2.21 The results in Exercise 2.20 support the sequential processing hypothesis.

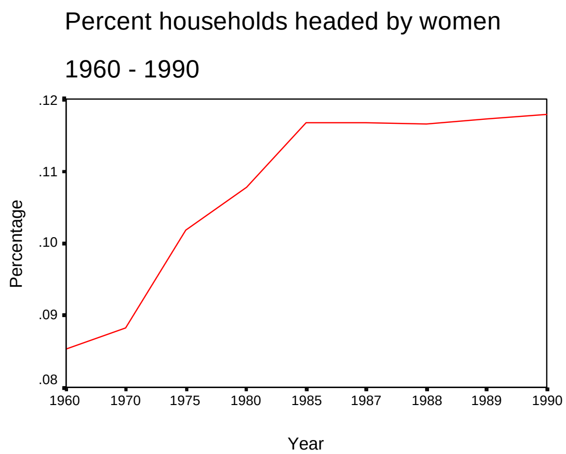
2.23 The data are not likely to be independent observations because the subject is probably learning the task over the early trials, and later getting tired as the task progresses. Thus responses closer in time are more likely to be similar than responses further away in time.

2.25 The amount of shock that a subject delivers to a white participant does not depend upon whether or not that subject has been insulted by the experimenter. On the other hand, black participants do suffer when the experimenter insults the subject.

2.27 AIDS cases among people aged 13—29 in U.S. population (in thousands)



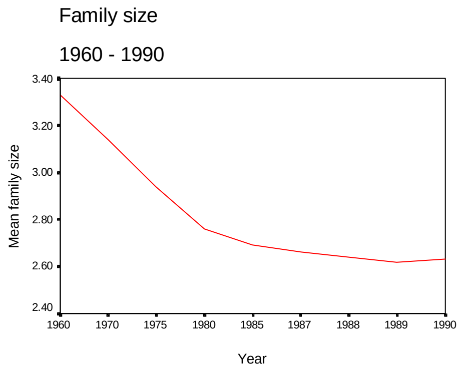
2.29 One way to look at these data is to plot the percentage of households headed by women and the family size separately against years. Notice that there is an uneven sampling of years.



.

a. There has been a dramatic increase in the percentage of households headed by women over the past 10 years.

b. There has also been a corresponding decrease in family size, part of which is perhaps due to the increase in single-parent families.



2.31 The mean falls above the median.

2.33 Rats running a straight alley maze:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{320}{15} = 21.33;\ \text{Median} = 21$$

2.35 Multiplying by a constant:

Original data (X):	8	3	5	5	6	2	$\overline{Y} = 4.83$ Median = 5 Mode = 5
Transformed data ($Y = 3X$)	24	9	15	15	18	6	$\overline{Y} = 14.5$ Median = 15 Mode = 15

$$3\overline{X} = \overline{Y}$$
$$3(4.83) = 14.5$$
$$14.5 = 14.5$$

$$3(\text{Med}_x) = \text{Med}_y$$
$$3(5) = 15$$
$$15 = 15$$

$$3(\text{Mo}_x) = \text{Mo}_y$$
$$3(5) = 15$$
$$15 = 15$$

2.37 Computer printout.

2.39 Computer exercise

2.41 For the data in Exercise 2.4:

$\text{range} = 17 - 1 = 16$

variance = $s_X^2 = \Sigma\left(X - \bar{X}\right)^2 = (10 - 10.2)^2 + (12 - 10.2)^2 + ... + (9 - 10.2)^2$

$= 11.592$

$\text{standard deviation} = s_X = \sqrt{s_X^2} = \sqrt{11.592} = 3.405$

2.43 For the data in Exercise 2.1:

The interval:

$\bar{X} \pm 2s_X = 18.9 \pm 2(4.496) = 18.9 \pm 8.992 = 9.908 \text{ to } 27.892$

From the frequency distribution in Exercise 2.1 we can see that all but two scores (31 and 40) fall in this interval, therefore 48/50 = 96% of the scores fall in this interval.

2.45

Original data: (reordered)	2	3	4	4	5	5	9	$\bar{X}_1 = 4.57$	$s_1 = 2.23$
$X_2 = X_1 + 3$	5	6	7	7	8	8	12	$\bar{X}_2 = 7.57$	$s_2 = 2.23$
$X_3 = X_1 - 2$	0	1	2	2	3	3	7	$\bar{X}_3 = 2.57$	$s_3 = 2.23$

$\bar{X}_1 = \frac{32}{7} = 4.57$ $\bar{X}_2 = \frac{53}{7} = \bar{X}_1 + 3$ $\bar{X}_3 = \frac{18}{7} = 2.57 = \bar{X}_1 - 2$

As we saw in Exercise 2.22, adding (or subtracting) a constant to (or from) a distribution adds (or subtracts) that constant from the mean of that distribution. Here we find that the standard deviation of that distribution is unchanged.

2.47 Original data: 5 8 3 8 6 9 9 7

$s_1 = 2.1$

If $X_2 = cX_1$, then $s_2 = cs_1$ and we want $s_2 = 1.00$

$s_2 = cs_1$

$1 = c(2.1)$

$c = 1/(2.1)$

Therefore we want to divide the original scores by 2.1

$X_2 = \frac{X_1}{2.1}$: 2.381 3.809 1.428 3.809 2.857 4.286 4.286 3.333

$s_2 = 1$

2.49 Boxplot for the data in Exercise 2.1 [Refer to data in Exercise 2.1 and cumulative distribution in Exercise 2.6]:

Median location = $(N + 1)/2 = 51/2 = 25.5$
Median = 18
Hinge location = (Median location + 1)/2 = $(25 + 1)/2 = 26/2 = 13$
Hinges = 17 and 20
H-spread = 20 - 17 = 3
Inner fences = Hinges $\pm$ 1.5*(H-spread)
= $17 – 1.5(3) = 17 – 4.5 = 12.5$
= $20 + 1.5(3) = 20 + 4.5 = 24.5$
Adjacent values = 15 and 24

```
0    5   10   15   20   25   30   35   40   45   50
+----+----+----+----+----+----+----+----+----+----+

          ***  --|| |----      *        *
```

2.51 Boxplot for ADDSC [Refer to stem-and-leaf in Exercise 2.15]:

Median location = $(N + 1)/2 = (88 + 1)/2 = 89/2 = 44.5$
Median = 50
Hinge location = (Median location + 1)/2 = $(44 + 1)/2 = 45/2 = 22.5$
Hinges = 44.5 and 60.5
H-spread = 60.5 - 44.5 = 16
Inner fences = Hinges $\pm$ 1.5*(H-spread)
= $60.5 + 1.5(16) = 60.5 + 24 = 84.5$
and = $44.5 – 1.5(16) = 44.5 – 24 = 20.5$
Adjacent values = 26 and 78

```
20        30        40        50        60        70        80        90
+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+

      ------------------|     |        |-------------------     **
```

2.53 Coefficient of variation for Appendix Data Set
$s/\bar{X} = 0.8614/2.456 = 0.351$

2.55 10% trimmed means of data in Table 2.6

3.13 3.17 3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34 3.36 3.38

Ten percent trimming would remove the two extreme observations at either end of the distribution, leaving

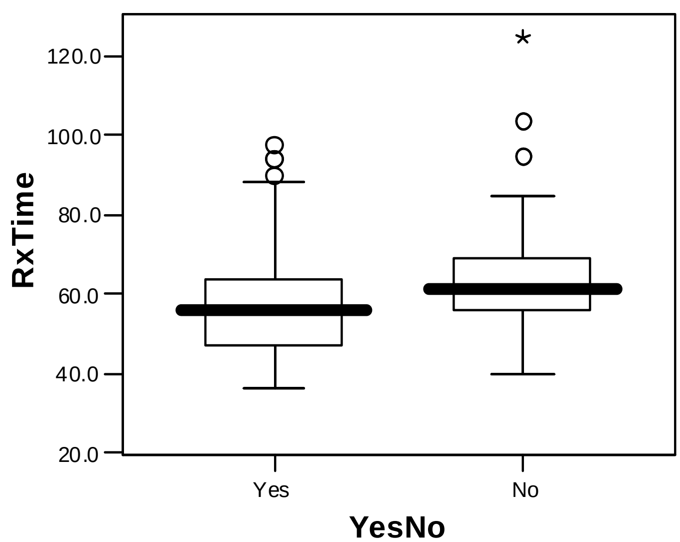
3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34

$$\bar{X} = \frac{52.28}{16} = 32.675$$

In this case the trimmed mean is very close to the untrimmed mean (3.266).

2.57 Reaction times when stimulus was present or absent.



2.59 This is an internet search that has no fixed answer.