1. (1.8a)

   (a) Refer to Table 1.4 on p. 13

Variable	**Classification**
AGE	Quantitative
SEX	Categorical
MANUF	Categorical
DIAG	Categorical
STAGE	Categorical
TOX	Categorical
DOSE	Quantitative
SCR	Quantitative
WEIGHT	Quantitative
GENERIC	Categorical

2. (1.10)

   (a) Ordinal

   (b) Quantitative

   (c) Quantitative

   (d) Quantitative

   (e) Quantitative

   (f) Categorical

3. (2.4)

   (a) Explanatory Variable: Cell phone use
   Response Variable: Occurrence of brain cancer

   (b) Explanatory Variable: Type of anesthetic
   Response Variable: Mortality rate or death

   (c) Explanatory Variable: Treatment group (i.e., vaccine or placebo)
   Response Variable: Poliomyelitis

4. (2.8) The first two eligible participants have IDs 04987 and 02927.

5. (2.12) This is **not** an SRS.

   This is because every nurse does not have an equal probability of being in the sample. An SRS requires that every conceivable combination of 9 nurses possibly end up in the sample. The current sample is restricted to 4 maternity nurses, 2 oncology nurses, and 3 surgical nurses. An SRS would allow for, say, 5 maternity nurses, no oncology nurses, and 4 surgical nurses. This problem describes a stratified random sample.

6. (3.2a,d)

   (a) Stemplot of hospitalization stay durations:

0	33344555567788999
1	0111147
2	
3	0

   × 10 days

   (d) Shape:

   - Positively skewed
   - Unimodal
   - One high outlier

   Location: The median = 8.

   Spread: Including the outlier, values vary from 3 to 30 days. Excluding the outlier, the stay lengths range from 3 to 17 days.

7. (3.6)

   (a) Stemplot of the waiting times at a public health clinic in minutes:

0	6
1	033669
2	23489
3	1235
4	259
5	15
6	3
7	227

   ×10 minutes

   Shape:

   - Bimodal
   - Asymmetrical

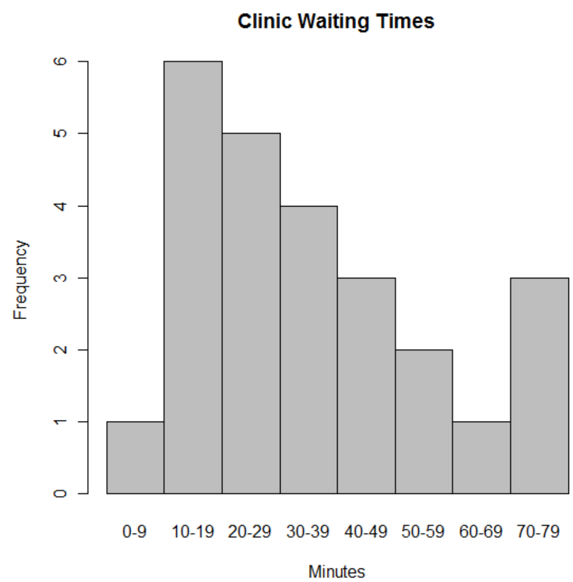
   Location: Median = 31

   Spread: Waiting times at this clinic range from 6 minutes to 77 minutes.

   (b) Frequency Table of Waiting Times, using class intervals of length 10. Note that the interval $0 - 9$ includes times greater than zero, and up to (but not including) 10.

Minutes	Frequency	Relative frequency (%)	Cumulative rel. frequency (%)
0 - 9	1	4%	4%
10 - 19	6	24	28
20 - 29	5	20	48
30 - 39	4	16	64
40 - 49	3	12	76
50 - 59	2	8	84
60 - 69	1	4	88
70 - 79	3	12	100%
	25	100%	–

28% (7 out of 25) of wait times were less than 20 minutes.
72% (18 out of 25) of wait times were at least 20 minutes.



8. (3.10) Stemplot of the durations of surgery in hours:

1	8
2	1345568
3	0115
4	
5	
6	5
7	0

$\times$ 1 axis multiplier

Shape:

- Unimodal
- Asymmetrical
- Positively skewed with two outliers: 6.5 hours and 7.0 hours.

Location: Median = 2.7 hours

Spread: Excluding the outliers, values range from 1.8 and 3.5 hours.

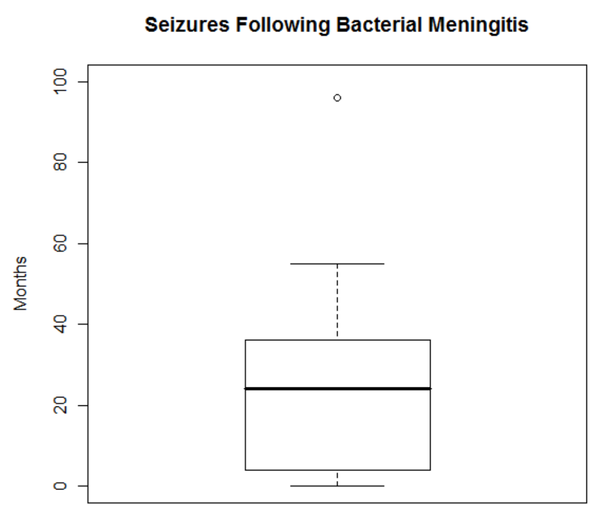
9. (4.4)

   (a) $\bar{x} = 25.9$
   Median= 24

   (b) The mean is higher than the median, suggesting that these data are asymmetrical and positively skewed.

   (c) Due to the positive skew, the median is a better measure of central location than the mean, since it is not strongly influenced by outliers.

10. (4.6) Boxplot of seizure onset after bacterial meningitis.



Yes, there is one upper outside value in this dataset (value=96).

Yes, there is evidence of asymmetry, with positive skew and one high outlier.

Note that to construct the boxplot, we follow the steps outlined in the Chapter 4 notes.

(1) Five-number Summary: {0.1, 4, 24, 36, 96}

(2) IQR = 36 - 4 = 32

Fence$_{\text{Lower}}$ $= 4 - (1.5)(32) = -44$. There are no outside values below the lower fence. The lower inside value is 0.10.

Fence$_{\text{Upper}}$ $= 36 + (1.5)(32) = 84$. There is one value above the upper fence (96). The largest value still inside the fence, i.e., the upper inside value, is 55.

(3) There are no outside values below the lower fence. There is one outside value above the upper fence (96).

(4) The lower inside value is 0.10. The largest value still inside the fence, i.e., the upper inside value, is 55.

11. (4.8) First we calculate $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(290) = 36.25$ and then we calculate s_1:

Data	Deviations	Squared deviations
68	$68 - 36.25 = 31.75$	$31.75^2 = 1008.0625$
22	$22 - 36.25 = -14.25$	$(-14.25)^2 = 203.0625$
36	$36 - 36.25 = -0.25$	$(-0.25)^2 = 0.0625$
32	$32 - 36.25 = -4.25$	$(-4.25)^2 = 18.0625$
42	$42 - 36.25 = 5.75$	$5.75^2 = 33.0625$
24	$24 - 36.25 = -12.25$	$(-12.25)^2 = 150.0625$
28	$28 - 36.25 = -8.25$	$(-8.25)^2 = 68.0625$
38	$38 - 36.25 = 1.75$	$1.75^2 = 3.0625$
$\sum x_i = 290$	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 1483.5$

$$\Longrightarrow \quad s_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i - \bar{x})^2\right]} = \sqrt{\frac{1}{7}(1483.5)} = \sqrt{211.9285} = 14.56\ \mu\text{g/m}^3$$

Recall that the standard deviation at site 2 is $s_2 = 2.9\ \mu\text{g/m}^3$. The numerical comparison of standard deviations (2.9 versus 14.6) confirms what we see on the side-by-side boxplots (Figure 4.5). Specifically, that the variability in the site 1 data is considerably greater than that of site 2.

12. (4.10)

$\bar{x}_{\text{Before}} = 70, \qquad \bar{x}_{\text{After}} = 70$

$s_{\text{Before}} = 14.14, \qquad s_{\text{After}} = 1.41$

No, we cannot tell if the drug worked just considering the means, since there are other measures of stability. From the calculation of the standard deviations, we see that variability decreased after treatment, indicating that the drug worked; that is, the drug effectively stabilized the individuals' heart rates. If we only look at the means, we might incorrectly conclude that the drug didn't work.

13. (4.14)

    (a) $\bar{x} = 25.91$
    $s = 27.37$

    (b) Median = 24
    IQR = $Q3 - Q1 = 36 - 4 = 32$

    (c) After removing the outlier,
    $\bar{x} = 20.07$
    $s = 18.26$
    Median = 18
    IQR = $Q3 - Q1 = 33.5 - 2.5 = 31.25$

    - After removing the outlier, the mean and standard deviation are much lower than before.
    - Removing the outlier lowers the median, but the IQR is almost exactly the same.

14. (4.18)

COPD		Controls
9	0	
8510	1	09
	2	034
	3	2

×10 axis multiplier

Group	n	$\bar{x}$	s
COPD	5	13.02	3.76
Controls	6	21.98	7.34