CHAPTER 2
Description of Samples and Populations

2.1.1 **(a)** i) Molar width
ii) Continuous variable
iii) A molar
iv) 36

(b) i) Birthweight, date of birth, and race
ii) Birthweight is continuous, date of birth is discrete (although one might say categorical and ordinal), and race is categorical
iii) A baby
iv) 65

• **2.1.2** **(a)** i) Height and weight
ii) Continuous variables
iii) A child
iv) 37

(b) i) Blood type and cholesterol level
ii) Blood type is categorical, cholesterol level is continuous
iii) A person
iv) 129

2.1.3 **(a)** i) Number of leaves
ii) Discrete variable
iii) A plant
iv) 25

(b) i) Number of seizures
ii) Discrete variable
iii) A patient
iv) 20

2.1.4 **(a)** i) Type of weather and number of parked cars
ii) Weather is categorical and ordinal, cars is discrete
iii) A day
iv) 18

(b) i) pH and sugar content
ii) Both variables are continuous
iii) A barrel of wine
iv) 7

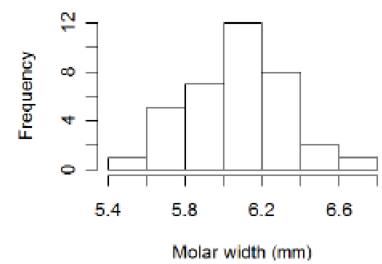
2.1.5 **(a)** i) body mass and sex
ii) body mass is continuous, sex is categorical
iii) A blue jay
iv) 123

(b) i) lifespan, thorax length, and percent of time spent sleeping

ii) Lifespan is discrete, thorax length and sleeping time are continuous
iii) A fruit fly
iv) 125

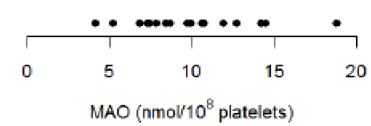
• **2.2.1** **(a)**There is no single correct answer. One possibility is:

Molar width	Frequency (no. specimens)
[5.4, 5.6)	1
[5.6, 5.8)	5
[5.8, 6.0)	7
[6.0, 6.2)	12
[6.2, 6.4)	8
[6.4, 6.6)	2
[6.6, 6.8)	1
Total	36



(b) The distribution is fairly symmetric.

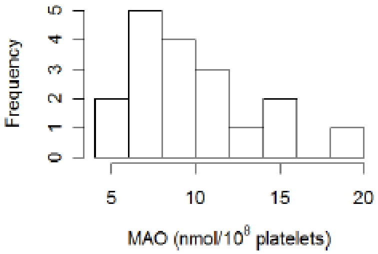
2.2.2



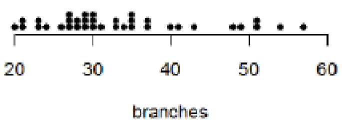
2.2.3 There is no single correct answer. One possibility is

MAO	Frequency (no. patients)
4.0-5.9	2
6.0-7.9	5
8.0-9.9	4
10.0-11.9	3
12.0-13.9	1
14.0-15.9	2

16.0-17.9	0
18.0-19.9	1
Total	18

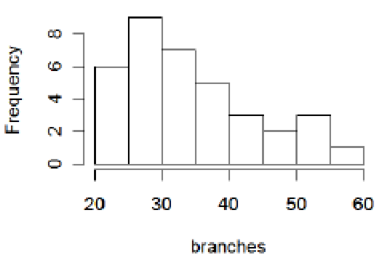


2.2.4



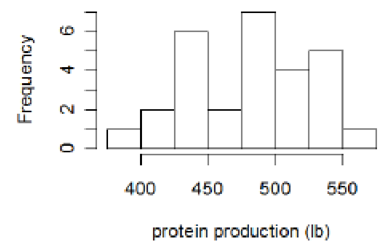
2.2.5 There is no single correct answer. One possibility is

Branches	Frequency (no. cells)
20-24	6
25-29	9
39-34	7
35-39	5
40-44	3
45-49	2
50-54	3
55-59	1
Total	36



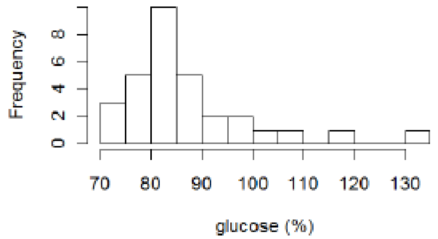
2.2.6 There is no single correct answer. One possibility is

Protein production	Frequency (no. cows)
375-399	1
400-424	2
425-449	6
450-474	2
475-499	7
500-524	4
525-549	5
550-574	1
Total	28

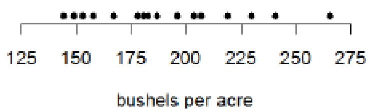


• **2.2.7** There is no single correct answer. One possibility is

Glucose (%)	Frequency (no. of dogs)
70-74	3
75-79	5
80-84	10
85-89	5
90-94	2
95-99	2
100-104	1
105-109	1
110-114	0
115-119	1
120-124	0
125-129	0
130-134	1
Total	31

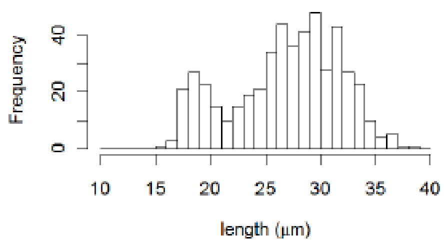


2.2.8 **(a)**



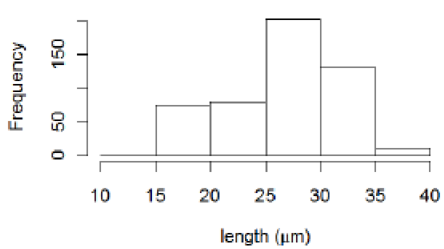
(b)
The distribution is very slightly skewed to the right.

2.2.9 **(a)**



(b) The distribution is bimodal.

(c) The histogram with only 6 classes obscures the bimodal nature of the distribution.



• **2.3.1** Any sample with $\Sigma y_i = 100$ would be a correct answer. For example: 18, 19, 20, 21, 22.

2.3.2 Any sample with $\Sigma y_i = 100$ and median 15 would be a correct answer. For example: 13, 14, 15, 28, 30.

2.3.3 $\bar{y} = \Sigma y_i/n = \frac{6.3 + 5.9 + 7.06.9 + 5.9}{5} = 6.40$ nmol/gm. The median is the 3rd largest value (i.e., the third observation in the *ordered* array of 5.9 5.9 6.3 6.9 7.0), so the median is 6.3 nmol/gm.

2.3.4 Yes, the data are consistent with the claim that the typical liver tissue concentration is 6.3 nmol/gm. The value of 6.3 fits comfortably near the center of the sample data.

• **2.3.5** $\bar{y} = 293.8$ mg/dl; median = 283 mg/dl.

• **2.3.6** $\bar{y} = 309$ mg/dl; median = 292 mg/dl.

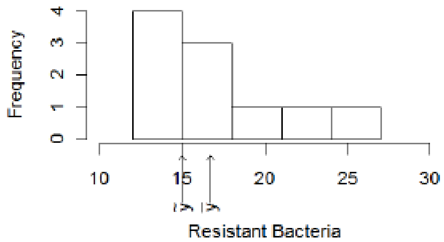
2.3.7 $\bar{y} = 3.492$ lb; median = 3.36 lb.

2.3.8 Yes, the data are consistent with the claim that, in general, steers gain 3.5 lb/day; the value of 3.5 fits comfortably near the center of the sample data. However, the data do not support the claim that 4.0 lb/day is the typical amount that steers gain. Both the mean and the median are less than 4.0; indeed, the maximum in the sample is less than 4.0.

2.3.9 $\bar{y} = 3.389$ lb; median = 3.335 lb.

2.3.10 There is no single correct answer. One possibility is

Resistant bacteria	Frequency (no. aliquots)
12-14	4
15-17	3
18-20	1
21-23	1
24-26	1
Total	10



(b) $\bar{y} = 16.7$, median $= \dfrac{15+15}{2} = 15$.

• **2.3.11** The median is the average of the 18th and 19th largest values. There are 18 values less than or equal to 10 and 18 values that are greater than or equal to 11. Thus, the median is

$\dfrac{10+11}{2} = 10.5$ piglets.

2.3.12 $\bar{y} = 375/36 = 10.4$.

• **2.3.13** The distribution is fairly symmetric so the mean and median are roughly equal. It appears that half of the distribution is below 50 and half is above 50. Thus, mean ≈ median ≈ 50.

2.3.14 Mean ≈ 35, median ≈ 40

2.4.1 (a) Putting the data in order, we have

13 13 14 14 15 15 16 20 21 26

The median is the average of observations 5 and 6 in the ordered list. Thus, the median is $\dfrac{15+15}{2} = 15$. The lower half of the distribution is

13 13 14 14 15

The median of this list is the 3rd largest value, which is 14. Thus, the first quartile of the distribution is $Q_1 = 14$. Likewise, the upper half of the distribution is

15 16 20 21 26

The median of this list is the 3rd largest value, which is 20. Thus, the third quartile of the distribution is $Q_3 = 20$.

(b) IQR $= Q_3 - Q_1 = 20 - 14 = 6$

(c) To be an outlier at the upper end of the distribution, an observation would have to be larger than $Q_3 + 1.5(\text{IQR}) = 20 + 1.5(6) = 20 + 9 = 29$, which is the upper fence.

• **2.4.2** (a) The median is the average of the 9th and 10th largest observations. The ordered list of the data is

4.1 5.2 6.8 7.3 7.4 7.8 7.8 8.4 8.7 9.7 9.9 10.6 10.7 11.9 12.7 14.2 14.5 18.8

Thus, the median is $\dfrac{8.7+9.7}{2} = 9.2$.

To find Q_1 we consider only the lower half of the data set:

4.1 5.2 6.8 7.3 7.4 7.8 7.8 8.4 8.7 9.7

Q_1 is the median of this half (i.e., the 5th largest value), which is 7.4.

To find Q_3 we consider only the upper half of the data set:

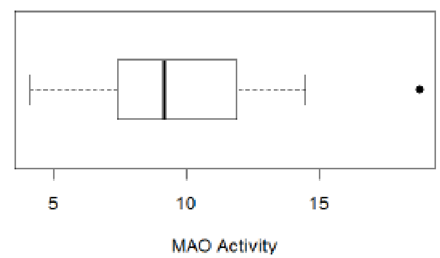
9.7 9.9 10.6 10.7 11.9 12.7 14.2 14.5 18.8.

Q_3 is the median of this half (i.e., the 5th largest value in this list), which is 11.9.

(b) IQR $= Q_3 - Q_1 = 11.9 - 7.4 = 4.5$.

(c) Upper fence $= Q_3 + 1.5 \times \text{IQR} = 11.9 + 6.75 = 18.65$.

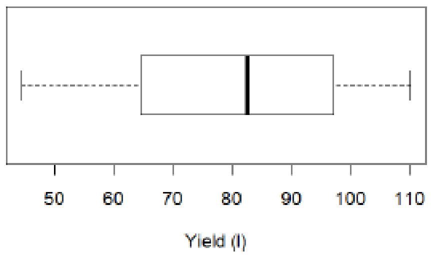
(d)



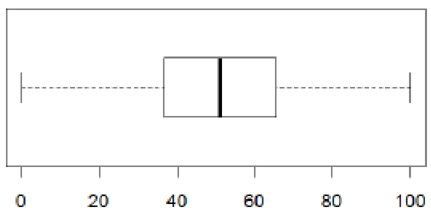
2.4.3 (a) Median = 82.6, $Q_1 = 63.7$, $Q_3 = 102.9$.

(b) IQR = 39.2.

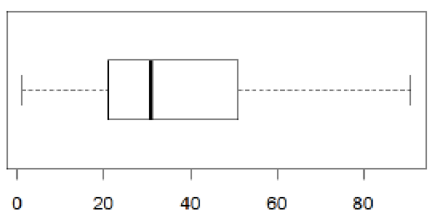
(c)



2.4.4 **(a)** $Q_1 = 35$, median = 50, $Q_3 = 65$.



(b) $Q_1 = 20$, median = 35, $Q_3 = 50$.



2.4.5 The histogram is centered at 40. The minimum of the distribution is near 25 and the maximum is near 65. Thus, boxplot (d) is the right choice.

2.4.6 Yes, it is possible that there is no data value of exactly 42. The first quartile does not need to equal a data value. For example, it could be that the first quartile is the average of two points that have the values 41 and 43.

2.4.7 **(a)** The IQR is 127.42 – 113.59 = 13.83.

(b) For a point to be an outlier it would have to be less than 113.59 – 1.5*13.83 = 92.845 or else greater than 127.42 + 1.5*13.83 = 148.165. But the minimum is 95.16 and the maximum is 145.11, so there are no outliers present.

2.4.8