Chapter 1.

Univariate data:
Classify, organise and display.

Situation: CARDIOVASCULAR DISEASE.

A doctor is to give a lecture on cardiovascular disease, i.e. diseases of the heart or blood vessels, including heart attack and stroke. Using statistics for the previous year she wants to start her lecture by saying:

"Did you know that of the ____ deaths in Australia last year, from all causes, ____ % of these were due to cardiovascular disease. This means that last year, on average, one Australian died from cardiovascular disease every ___ minutes."

Copy and complete the above introduction to her talk using the information given in the table below to fill in the blanks. (Give the percentage to the nearest percent and the time to the nearest minute.)

TOTAL AUSTRALIAN DEATHS FOR YEAR PRIOR TO TALK: All ages

CAUSE OF DEATH	MALE	FEMALE	PERSONS
CARDIOVASCULAR DISEASE			
Coronary heart disease	12 433	11 137	23 570
Stroke	4 668	6 845	11 513
Other cardiovascular	4 856	6 195	11 051
(Sub-total)	21 957	24 177	46 134
CANCERS	22 039	17 183	39 222
TRANSPORT ACCIDENTS	1 224	414	1 638
ALL OTHER	22 021	21 699	43 720
All causes	67 241	63 473	130 714

[Source of data: National Heart Foundation of Australia and The Australian Bureau of Statistics.]

The situation on the previous page involved you in making sense of information, or **data**, that was presented as a table, extracting relevant information from that table, and summarising that data in terms of percentages and times.
The data involved in the table would have been collected from government data bases where the cause of death for each deceased person, as stated on the death certificate, would be stored.

Data collection is often carried out to investigate some aspect of our lives. The methods by which we collect that data can vary as can the types of data we collect. The initial chapters of this text consider

the various types of data,
organising and displaying data,
describing and interpreting data,
comparing sets of data.

Types of data.

For some investigations we collect the data ourselves by asking questions, by measuring, by experiment etc. This is called **primary data** - data we have collected ourselves.

Sometimes it is appropriate to use the data already collected by others, as in the situation on the previous page. For us this would be **secondary data** - data collected by others.

Note: Internet access provides us with a ready supply of information regarding all sorts of subjects. Whilst much of this may well be valid we need to be cautious before assuming certain things about the information. We might consider such things as:

Was the method of data collection appropriate?
Is the information fairly presented?
Is the data collected from everyone in a particular situation or was sampling involved?
Are any summary statements correct?
Are any conclusions reasonable?

One way to have confidence in the validity of the information is to use data from a reputable source. The *Australian Bureau of Statistics* (ABS), for example, would be one such credible source, as would other Government departments. Such sources are likely to have correctly considered collection methods, suitable presentation of data, appropriate conclusions etc.

Variables.

If we ask someone what their favourite colour is or how tall they are the answers we get will vary from one person to another. Not everyone has blue as their favourite colour, not everyone is 163 cm tall etc. The responses will vary. Favourite colour and height are examples of **variables**.

If we are considering <u>one</u> variable, for example favourite colour, any data we collect will be **univariate**. However if we wanted to investigate how favourite colour may change with the age of a person we would be considering two variables – favourite colour and age. In such cases we would be considering **bivariate** data. This unit will only consider univariate data. Unit three of the *Mathematics Applications* course considers bivariate data.

Categorical variables.

Consider the following questions that might be asked in some data collection activities:

- How do you get to school (or work) - walk, cycle, car, bus, train, other?
- Are you male or female?
- What is your favourite colour?
- Have you ever lived in a country other than Australia?
- How would you rank your fitness level: Low, Medium or High?
- What would you consider the best description of the current outside temperature - cold, cool, warm, or hot?
- What is your house number?
- What number are you in your rugby team?

Each of these questions allow us to place the person who responds into a particular **category** or **group**. The response might allow us to place the respondent in the group of males, or the group of people whose favourite colour is red, those who regard their fitness level as high, or who live in houses numbered eleven, or play number eight in their rugby team etc. The variable concerned, be it mode of transport, gender, favourite colour etc are all examples of **categorical variables.** Data associated with a categorical variable is called **categorical data.**

Notice that the first four questions each involve categories for which order is irrelevant. In general it would be pointless to suggest an order to the modes of transport, or to the gender categories, or place red ahead of blue as a category. We may rank order according to the numbers of people in that category but not on the basis of the category *names* themselves which have no natural order about them. Such categorical variables are called **nominal categorical variables.** (Name - nominal).

However the last four questions each involve categories that do have a natural *order* about them. These last four questions involve **ordinal categorical variables,** i.e. variables that do have a natural order about them. (Order - ordinal)

Notice though that whilst categorical variables can have categories that have numbers assigned to them these numbers are simply labels, they have no numerical significance, as in the responses to the last two questions in the above list.
Houses numbered 11, for example, are not necessarily bigger, better or more expensive than houses numbered 10. The house number is simply a label.
The number 8 player in one rugby team is not necessarily bigger, faster or better than the number 6 in another team. The number is simply a label indicating a position played and has no numerical significance.

Categorical variables.

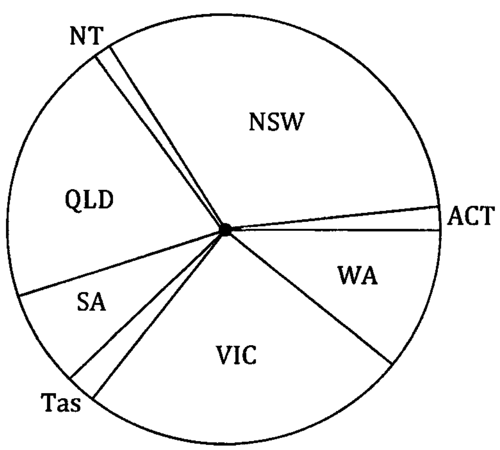
Nominal	**Ordinal**
Based purely on the names of the various categories no order is suggested.	The names of the categories, even though only labels, do suggest an order.

Displaying Categorical data.

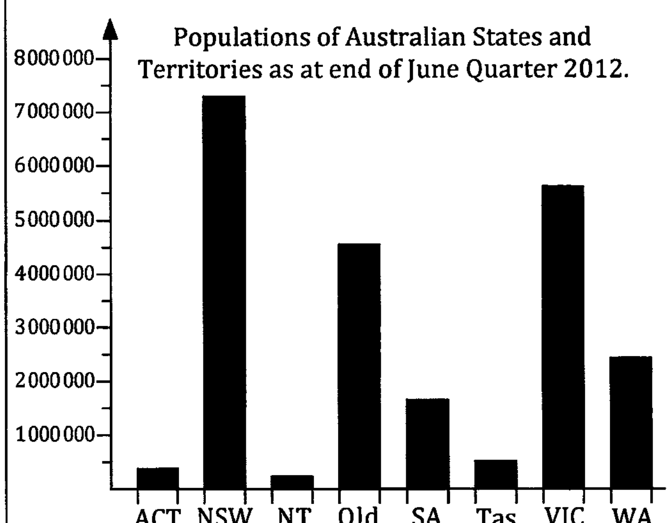
We tend to use tables, pie charts and bar graphs to organise and display categorical data, as shown below.

State or Territory		Population at end of June quarter 2012 (nearest hundred)
Australian Capital Territory	(ACT)	374 700
New South Wales	(NSW)	7 290 300
Northern Territory	(NT)	234 800
Queensland	(Qld)	4 560 100
South Australia	(SA)	1 654 800
Tasmania	(Tas)	512 000
Victoria	(Vic)	5 623 500
Western Australia	(WA)	2 430 300
Total		22 680 500

Source: Australian Bureau of Statistics.



Pie Chart of Australian State and Territory populations as at end of June Quarter 2012.
(Total population 22 680 500.)



Notice that we may choose some particular order in which to display this nominal categorical data, for example in alphabetical order (as used above), or perhaps in order of population size or land area, or perhaps in order of state or territory formation date, or in order of number of letters in their non abbreviated titles (!), etc, but on the basis of the categories themselves they have no natural order about them.

Exercise 1A

For 1 to 12 classify each of the categorical variables given as either *nominal* or *ordinal.*

1. Country of birth.
2. Preferred writing hand: Left, Right or no preference.
3. Playing number in a basketball team.
4. Home telephone number.
5. Blood group: A, B, A/B or O.
6. The number of the bus used to get to school.
7. A rating on sporting ability: Low, medium or high.
8. The construction type of a house: Brick, concrete, steel, timber, or other.
9. The state or territory of Australia that a person lives in.
10. Ow zumwon rayts ther spellin ablity: Poor, Okay, Good, Very Good, Excellent.
11. How someone voted in the last election.
12. The type of crop grown: Wheat, Barley, Oats etc

13. OFFENCES AGAINST PROPERTY.

    In one year in Western Australia there were 197117 offences categorised as an *offence against property*. These offences were further categorised as one of:

    - Burglary 38410 offences
    - Theft 81724 offences
    - Fraud 8552 offences
    - Graffiti 13762 offences
    - Stealing motor vehicle 7618 offences
    - Receiving 1547 offences
    - Arson 1269 offences
    - Property damage 44235 offences

    [Source of data: Western Australia Police.]

    Display this information both as a bar chart and as a pie chart and comment on the advantages and disadvantages of each form of display.

14. HOSPITAL BEDS

    In one particular year the number of hospital beds available for patient care in each state or territory of Australia were as follows:

NSW	Vic	Qld	WA	SA	Tas	NT	Act
27 543	22 502	16 623	8 138	8 021	2 737	764	1 083

    The fact that NSW had the largest number of beds available should come as no surprise because it had the largest population. To be able to compare these numbers more meaningfully we need to consider them with respect to the population of each state or territory which, in that particular year was as follows:

NSW	Vic	Qld	WA	SA	Tas	NT	Act
5 902 925	4 417 821	2 966 696	1 637 072	1 447 118	467 388	166 823	289 344

    [Based on data from the Australian Bureau of Statistics.]

    For each state or territory calculate the number of beds per 10000 of population for this particular year, giving your answers correct to the nearest whole number, and display your answers as a bar graph.

    Research: The above figures were actually for the early 1990s. Try to find more recent data for the number of beds per 10000 of population in each state and territory and comment on how it compares with the data given above.

Numerical variables.

Consider the following questions that might be asked in some data collection activities:

How many people live in your house?
How many people in your Mathematics class?
How many pets do you have?
How many of your teeth have been filled in some way?
How tall are you?
How far have you walked today?
What is your blood pressure reading?
What weight are you?

The responses to each of these questions will be **numerical** and the number will not just be a label, it will indicate a size or an amount. Some form of counting or measuring will be required to be done, or to have been done, for the response to be given. In each case the variable concerned, be it the number of people in a house, how many pets you have, your height, your blood pressure reading etc, are all examples of **numerical variables.** Data associated with a numerical variable is called **numerical data.**

Notice that the first four questions each involve responses that can only take specific values, in this case integer values. We cannot have 2·3 people living in a house, we cannot have 28·6 people in a Mathematics class. Numerical variables that can only take integer values are called **discrete variables.** Data associated with discrete variables is **discrete data.**

However the last four questions each involve responses that can take any value (usually within some realistic range). We can be 164·5 cm tall, we can walk 5·278 km in a day. Responses no longer have to be integer values and in practice the limit on the values taken are those of reasonableness (eg we cannot have 18·24 metres for the height of someone) and the accuracy of the measurement instrument used. Numerical variables that can take any value in an interval are called **continuous variables.** Data associated with a continuous variable is **continuous data.**

Generally, if counting is involved we have a discrete variable, if measurement is involved we have a continuous variable.

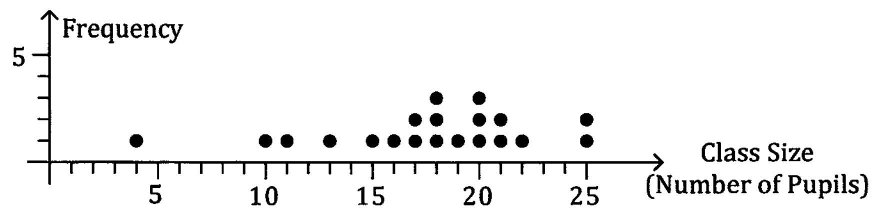
Exercise 1B

For 1 to 12 classify each of the numerical variables given as either *discrete* or *continuous.*

1. The number of rooms in a house.
2. The area of the block of land a house stands on.
3. The number of brothers and sisters a person has.
4. The length of a person's handspan.
5. The time it takes to get to school.
6. The lifetime of a rechargeable battery before it needs recharging.
7. The number of car thefts in Australia in a week.
8. The temperature in degrees Celsius.
9. The number of people visiting a supermarket in a day.
10. The length of a car.
11. The capacity of a swimming pool.
12. The weight of a frog.

Displaying numerical data.

In an attempt to analyse the use of its teachers a school noted the number of students in each of 20 year eleven classes. The school did not allow classes to run with more than 25 students and the smallest group had just 4 students. The number of students in the 20 classes gave rise to the following **dot frequency diagram**:

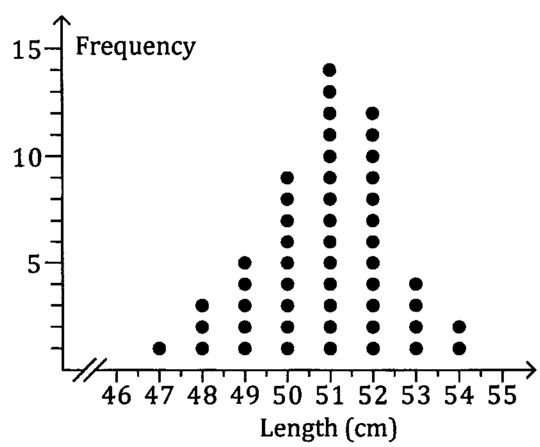


The variable quantity here, i.e. the number of students in each class, is an example of a **discrete** variable because the numerical variable can only take particular values, in this case integer values, from the given low of 4 to the high of 25.

Consider instead the continuous variable of the length of a new born baby and suppose that the lengths of 50 babies, recorded to the nearest centimetre, gave rise to the following table:

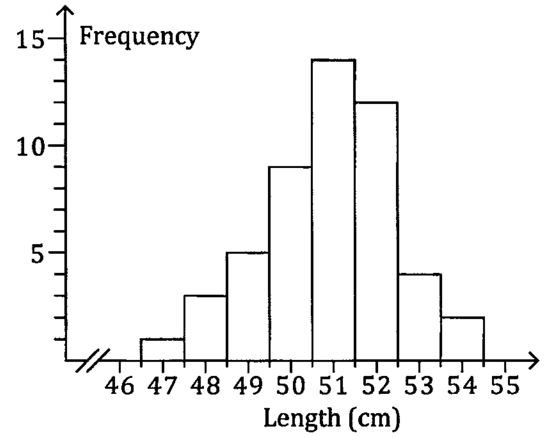
Length (cm)	47	48	49	50	51	52	53	54
Frequency	1	3	5	9	14	12	4	2

We could display this information as a dot frequency diagram, as shown below.



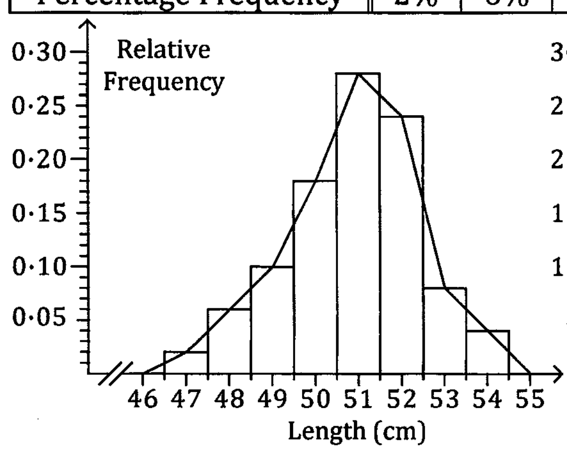
However it would be better to acknowledge the continuous nature of the data. The two babies recorded as 54 cm in length could have been anywhere from 53·5 cm to 54·5 cm. We can show this if we display the data as a **frequency histogram**. This is similar to a bar graph but always has frequency on the vertical axis, an ordered numerical scale on the horizontal axis and no gaps between the bars. Such a histogram for the data is shown on the next page.

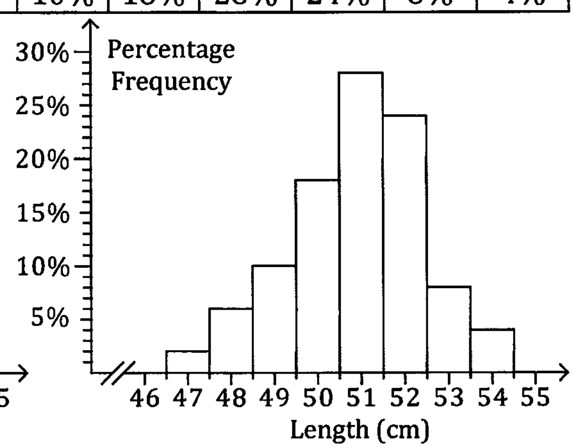
There was 1 baby recorded as 47 cm, when measured to the nearest centimetre. Thus for the 47 cm class the lower class boundary is 46·5 cm and the upper class boundary is 47·5 cm. This gives the first column in the histogram which is shown completed below.



Note
- A histogram does not have gaps between the columns because, with continuous data, each class begins where the previous one leaves off. (Except by the apparent "gap" between columns if a column has a frequency of zero.)
- We would usually have about 6 to 10 intervals. Too many and the table can be unmanageable, too few causes unnecessary bunching.
- It is usual to have class intervals of equal width. Intervals of different width could be misleading (and will be avoided in this text).
- The vertical axis on a histogram is always frequency.
- The horizontal axis on a histogram is a number line.
- Sometimes the frequency in each class may be shown as a fraction of the whole (*relative frequency*) or as a percentage of the whole (*percentage frequency*). Note that one of the histograms below also shows a **frequency polygon**, formed by connecting the middle of the top of each bar to the next, thus making the overall shape and continuity more evident.

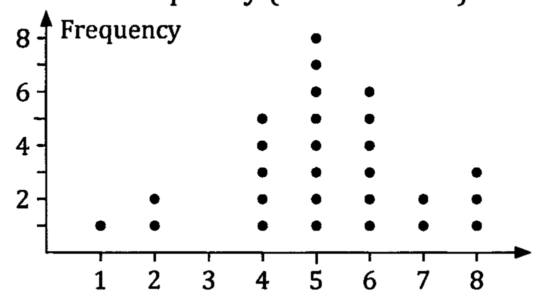
Length (cm)	47	48	49	50	51	52	53	54
Frequency	1	3	5	9	14	12	4	2
Relative Frequency	0·02	0·06	0·10	0·18	0·28	0·24	0·08	0·04
Percentage Frequency	2%	6%	10%	18%	28%	24%	8%	4%



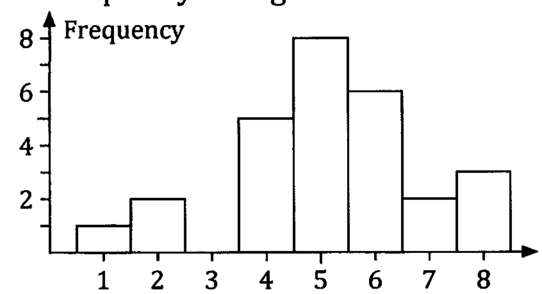


- Whilst we might expect that discrete data would not be shown as a histogram because there would be gaps between the columns, histograms are such a convenient form of representation that they are frequently used to display discrete data. In such cases we choose our class boundaries to be mid way between the possible discrete values. For example the discrete data shown below left as a dot frequency diagram can be displayed as a frequency histogram, as shown below right.

Dot frequency (discrete data).



Frequency histogram of the data.



- Histograms can be a useful form of display when we have discrete data that is *grouped* for convenience. For example, consider the following 25 scores:

35	46	12	34	18	20	25	24	11
14	29	9	27	23	32	38	30	17
22	19	36	28	33	4	21		

No scores are repeated so if we were to display the scores as a frequency table, or as a dot frequency graph, we would have twenty five scores shown, each with a frequency of 1. In this case the data may be better presented grouped. Using the intervals 1 - 5, 6 - 10, 11 - 15, 16- 20 etc., the grouping becomes:

Score	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41- 45	46-50
Frequency	1	1	3	4	5	4	4	2	0	1

Some information is now lost because the 25 scores themselves are no longer given but the grouping can make the overall distribution more evident.
The grouped data could be displayed as a histogram.

Histograms and bar charts.

Both bar charts and histograms can show the frequency of something occurring, so what is the difference between them? One important difference is that histograms show a normal number line on the horizontal axis, bar graphs show categories. This means that the bars of a bar graph are not bound by order and can be moved around. We might for example arrange the bars in order of increasing height. The bars in a histogram cannot be moved around. They must be presented in order, with the horizontal axis giving this order. Further, the beginning of one category in categorical data does not logically take over from the end of another. If we are drawing a bar graph about the pets people have, our categories might be cat, dog, horse, etc. The categories are quite separate - a dog does not start where a cat leaves off. Hence bar charts tend to have gaps between the bars. However with ordered numerical data, especially of a continuous nature, one number interval does indeed take over where the previous one left off. Hence we put no gaps between the bars of a histogram.

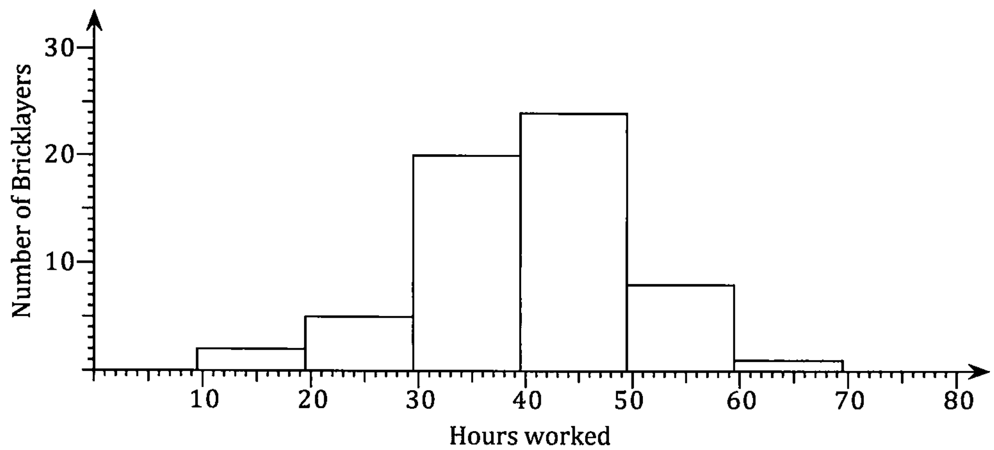
Example 1

Sixty self employed bricklayers were surveyed regarding the number of hours they worked during one particular week (to the nearest hour). The results of the survey are shown below.

Number of Hours	Frequency (i.e. N$^{o.}$ of bricklayers)
10 → 19	2
20 → 29	5
30 → 39	20
40 → 49	24
50 → 59	8
60 → 69	1

Display the data as a histogram.

The first column of the histogram will extend from 9·5 to 19·5, the second column from 19·5 to 29·5 and so on. The completed histogram is shown below.



Example 2.

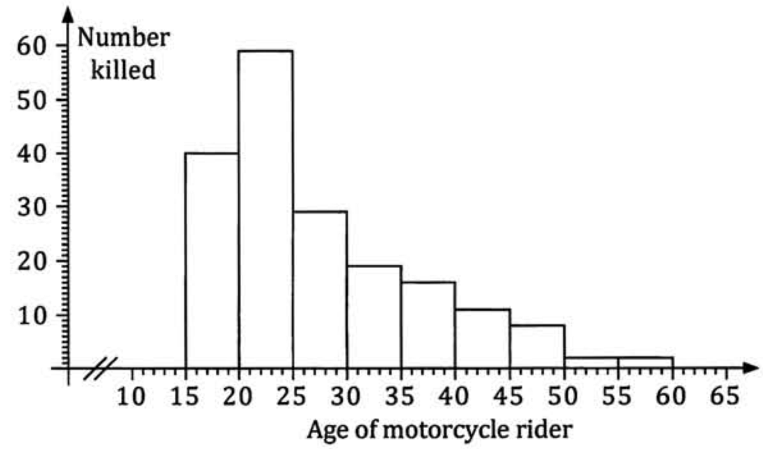
The road accident statistics for a country for one year showed that for motorcyclists (drivers not passengers) in the age range fifteen to fifty-nine, 186 had died in road accidents with the distribution of the ages of these riders as shown on the right,

Display this information as a frequency histogram.

Age (x yrs)	Drivers Killed
$15 \leq x < 20$	40
$20 \leq x < 25$	59
$25 \leq x < 30$	29
$30 \leq x < 35$	19
$35 \leq x < 40$	16
$40 \leq x < 45$	11
$45 \leq x < 50$	8
$50 \leq x < 55$	2
$55 \leq x < 60$	2

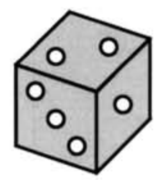
The first column of the histogram will extend from 15 to 20, the second column from 20 to 25 and so on.

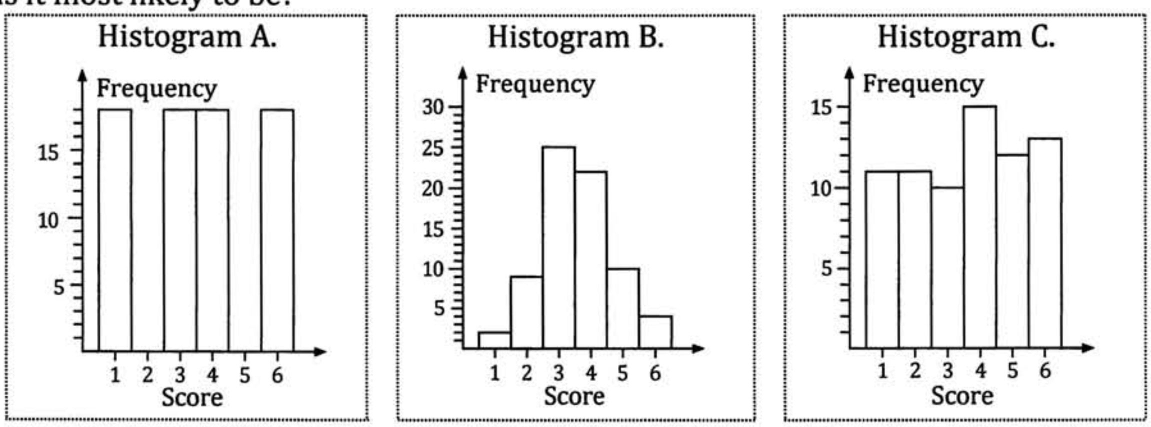
The completed histogram is shown on the right.



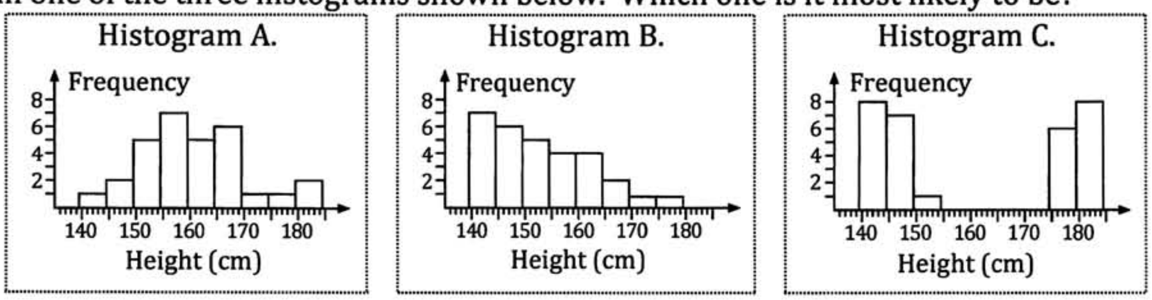
Exercise 1C

1. A normal, fair, six sided die is rolled 72 times and the number displayed on the uppermost face is noted each time.
Given that one of the three frequency histograms shown below displays the results obtained for these 72 rolls which of the three is it most likely to be?

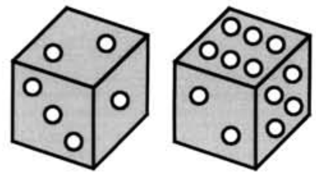


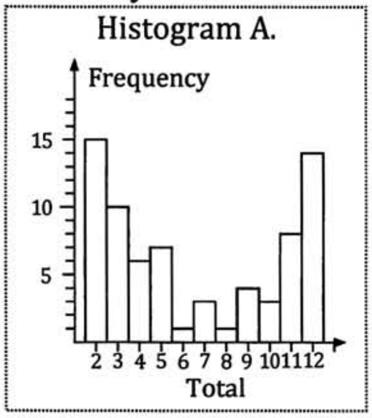


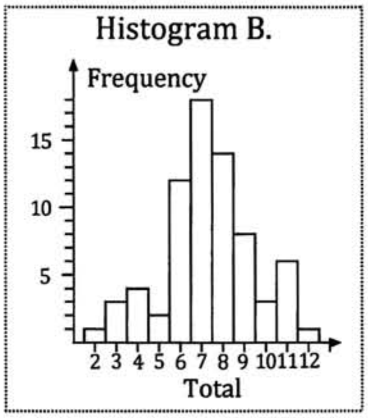
2. The thirty students in a class measured their heights and the results are displayed in one of the three histograms shown below. Which one is it most likely to be?

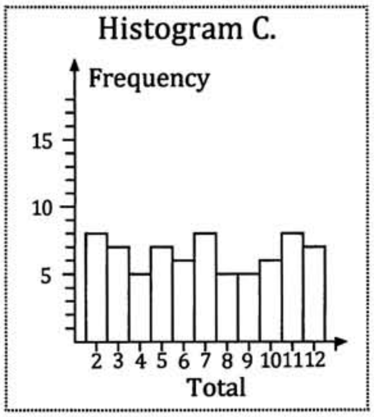


3. Two normal fair six sided dice are rolled and the numbers displayed on the two uppermost faces are added together and the total noted.
This process is carried out 72 times.
Given that one of the three frequency histograms shown below displays the results obtained for this activity which of the three is it most likely to be?









4. What shape will the histogram be?
Sketch what you consider to be a reasonable histogram for each of the following. Your sketch should show what you consider to be a reasonable shape for the histogram to have. There is no need to include any numbers on the axes of your sketch graphs except for parts (a) and (b) which should have numbers on the horizontal axis.
   (a) Rolling a fair eight sided die approximately one hundred times.
   (b) The number of children in approximately one hundred randomly selected families, each of which have at least one child.
   (c) The heights of a large number of adult males.
   (d) The straight line distance from home to school for students at your school.
   (e) The lengths of new born babies.

For at least some of the following questions generate the required histogram using a computer or calculator.

5. HORTICULTURE.
Some seeds were planted and, some weeks later, the heights of the seedlings were measured and recorded, to the nearest centimetre. The results were as follows:

Length (cm)	6	7	8	9	10	11	12	13	14	15
Frequency	2	4	10	15	14	11	9	7	3	1

Display this information as a frequency histogram.

6. METEOROLOGY.
The maximum temperature recorded at Perth airport for each day of December in one particular year gave rise to the following data:

Maximum Temperature (° C)	20 → 24	25 → 29	30 → 34	35 → 39	40 → 44
Frequency (number of days).	3	13	8	6	1

[Source: Bureau of Meteorology.]

Display this information as a frequency histogram.

7. ROAD ACCIDENTS.
The road accident statistics for a country for one year showed that for drivers in the age range fifteen to fifty-nine, 708 had died in road accidents. The distribution of the ages of these drivers was as follows:

Age (x yrs)	Drivers Killed
$15 \leq x < 20$	138
$20 \leq x < 25$	131
$25 \leq x < 30$	75
$30 \leq x < 35$	95
$35 \leq x < 40$	79
$40 \leq x < 45$	71
$45 \leq x < 50$	57
$50 \leq x < 55$	39
$55 \leq x < 60$	23
Total	708

Display this information as a frequency histogram.

8. TIME ESTIMATION.
One hundred and fifty students were asked to estimate a time period of one minute. The time periods they thought were one minute were actually the following number of seconds, to the nearest second:

31	68	46	66	54	48	70	60	62	48	97	53	50	56	60
52	56	92	50	43	65	45	80	53	64	56	67	59	41	49
65	75	50	51	66	75	50	56	40	57	64	44	69	71	51
51	64	89	74	49	54	57	67	54	59	47	79	51	54	50
59	90	49	61	52	64	77	46	74	48	66	49	76	66	41
43	50	81	62	68	44	49	66	52	84	45	84	42	52	59
65	74	42	73	54	50	73	60	49	60	54	52	69	56	50
62	47	51	45	50	67	59	38	65	46	56	85	48	54	51
50	67	65	54	65	48	51	54	62	52	51	53	70	43	57
47	64	54	69	43	86	62	69	51	64	52	76	64	68	46

Arrange the data as a grouped frequency table with classes of equal width as follows: 30 - 39, 40 - 49, etc., up to 90 - 99.
Display the grouped data as a frequency histogram.

9. WEIGHT.
Two hundred males and two hundred females, all aged between 30 and 40, took part in a survey which recorded (amongst other things) the weight of each person, recorded to the nearest kg. The results were as follows:

MALES	
Weight (kg)	Frequency
30 - 39	0
40 - 49	2
50 - 59	10
60 - 69	35
70 - 79	68
80 - 89	53
90 - 99	21
100 - 109	9
110 - 119	2
Total	200

FEMALES	
Weight (kg)	Frequency
30 - 39	5
40 - 49	12
50 - 59	73
60 - 69	66
70 - 79	27
80 - 89	8
90 - 99	5
100 - 109	3
110 - 119	1
Total	200

Display these results as two separate frequency histograms and include the frequency polygon on each one.

10. HEIGHT.
Fifty males and fifty females, all aged between 20 and 30, took part in a survey which recorded (amongst other things) the height of each person, recorded to the nearest cm. The results were as follows:

MALES	
Height (cm)	Frequency
140 - 149	0
150 - 159	1
160 - 169	6
170 - 179	26
180 - 189	14
190 - 199	2
200 - 209	1

FEMALES	
Height (cm)	Frequency
140 - 149	1
150 - 159	11
160 - 169	28
170 - 179	8
180 - 189	2
190 - 199	0
200 - 209	0

Display these results as two separate **percentage** frequency histograms.

Miscellaneous Exercise One.

This miscellaneous exercise may include questions involving the work of this chapter and the ideas mentioned in the preliminary work section at the beginning of the book.

1. State true or false for each of the following statements:

(a)	$3^2 = 6$	(b)	$(-3)^2 = -9$
(c)	$-5 - 3 = 8$	(d)	$5^2 + 3^2 = 8^2$
(e)	$5 + 3 \times 2 = 16$	(f)	$3(x + 1) = 3x + 1$

2. Classify each of the variables mentioned in parts (a) to (h) as one of:

Nominal categorical	Ordinal categorical
Discrete numerical	Continuous numerical

(a) Favourite soccer team.
(b) Waist measurement.
(c) Number of people in a car.
(d) Mode of transport: Walk, cycle, bus, train, tram, other.
(e) Interest in sport: Not at all, weak, medium, strong, full on.
(f) Nationality of mother.
(g) Distance from home to school.
(h) The number of peas in a pod.

3. Find the mean of each of the following sets:

(a) 131, 120, 141, 122, 136.
(b) 2·4, 3·7, 1·9, 0, 2·3, 3·2, 1·6, 1·7.
(c) 27, 18, 31, 33, 39, 27, 41, 29, 21, 27.

4. Find the median of each of the following sets:

(a) 15, 17, 21, 22, 23, 25, 25.
(b) 19, 21, 13, 28, 22, 25, 19, 22, 17.
(c) 10, 17, 11, 15, 23, 11, 21, 12, 17, 9.

5. Find the mode of each of the following sets:

(a) 3, 2, 1, 4, 3, 0.
(b) 11, 13, 17, 10, 13, 14, 17.
(c) 22, 15, 21, 22, 18, 19, 16, 22, 17, 19.

6. The instructions for mixing a weedkiller says to mix concentrate and water in the ratio 1 : 300.
How much water should be added to 25 millilitres of concentrate?

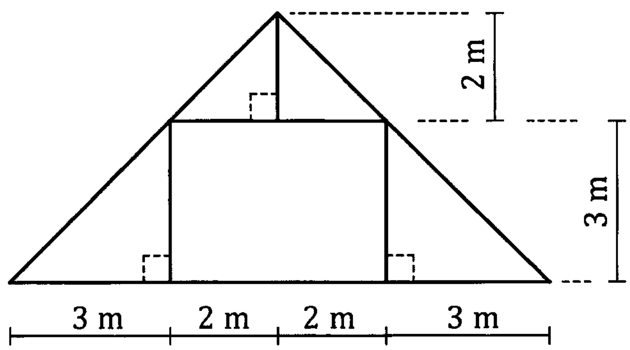
7. Find the mean, median, mode and range of each of the following sets:

(a) 33, 37, 38, 40, 40.
(b) 131, 93, 124, 107, 68, 131, 70, 110, 84.
(c) 18, 15, 17, 18, 15, 18, 18, 17, 19, 17.

8. Expand each of the following and simplify where possible.

(a) $3(5x - 2)$	(b) $4(7 - 2x)$	(c) $-3(2x + 7)$
(d) $8(1 - 2x)$	(e) $5(2p - 7)$	(f) $3(2h - 5)$
(g) $5 + 2(1 + 3x)$	(h) $4(2x + 1) - 5(3 + 2x)$	(i) $2x + 3(5 - 2x)$
(j) $2(5 + q) - 3(1 - 2q)$	(k) $6(2w + 3) - 5w + 4$	(l) $2(p + 6) - 4(3 - p)$

9. The framework shown on the right is to be made of steel. Find the total length of steel required for the framework giving your answer rounded up to the next whole metre.



10. HOLIDAY ACTIVITIES.

A sports club offers holiday activities for young people aged at least 5 but under 15. The enrolment forms required participants to give, amongst other things, their age in years and months. To determine which ages their programme of activities suited, club officials considered the ages of participants. These are shown below.

5y 3m	14y 5m	7y 2m	6y 2m	10y 5m	8y 10m	8y 3m	7y 5m
12y 2m	6y 1m	13y 3m	9y 9m	6y 1m	6y 3m	11y 5m	8y 5m
9y 11m	13y 0m	7y 10m	12y 2m	8y 0m	14y 7m	6y 9m	9y 11m
8y 7m	7y 1m	8y 0m	7y 5m	7y 0m	5y 3m	10y 0m	6y 7m
14y 10m	11y 2m	5y 8m	12y 1m	11y 7m	7y 3m	12y 3m	6y 11m
6y 1m	7y 9m	10y 2m	11y 2m	5y 2m	7y 3m	7y 4m	9y 2m
13y 2m	12y 4m	7y 2m	6y 3m	9y 4m	8y 0m	6y 9m	7y 7m
11y 11m	6y 1m	13y 9m	13y 7m	6y 9m	12y 7m	11y 2m	6y 7m
5y 1m	9y 0m	7y 7m	14y 0m	13y 7m	6y 9m	9y 7m	12y 11m
8y 7m	14y 1m	13y 9m	5y 11m	8y 11m	12y 7m	5y 7m	8y 1m
8y 6m	7y 11m	6y 4m	13y 10m	7y 6m	6y 5m	6y 2m	6y 6m

(a) Arrange these ages into a grouped frequency table as follows (tally shows first column entered):

Age (x years)	Tally	Frequency
$5 \le x < 6$	//	
$6 \le x < 7$	/	
$7 \le x < 8$		
$8 \le x < 9$	///	
$9 \le x < 10$	/	
$10 \le x < 11$		
$11 \le x < 12$	/	
$12 \le x < 13$	/	
$13 \le x < 14$	/	
$14 \le x < 15$	/	

(b) Display the grouped data as a frequency histogram.