CHAPTER 12 Data Handling

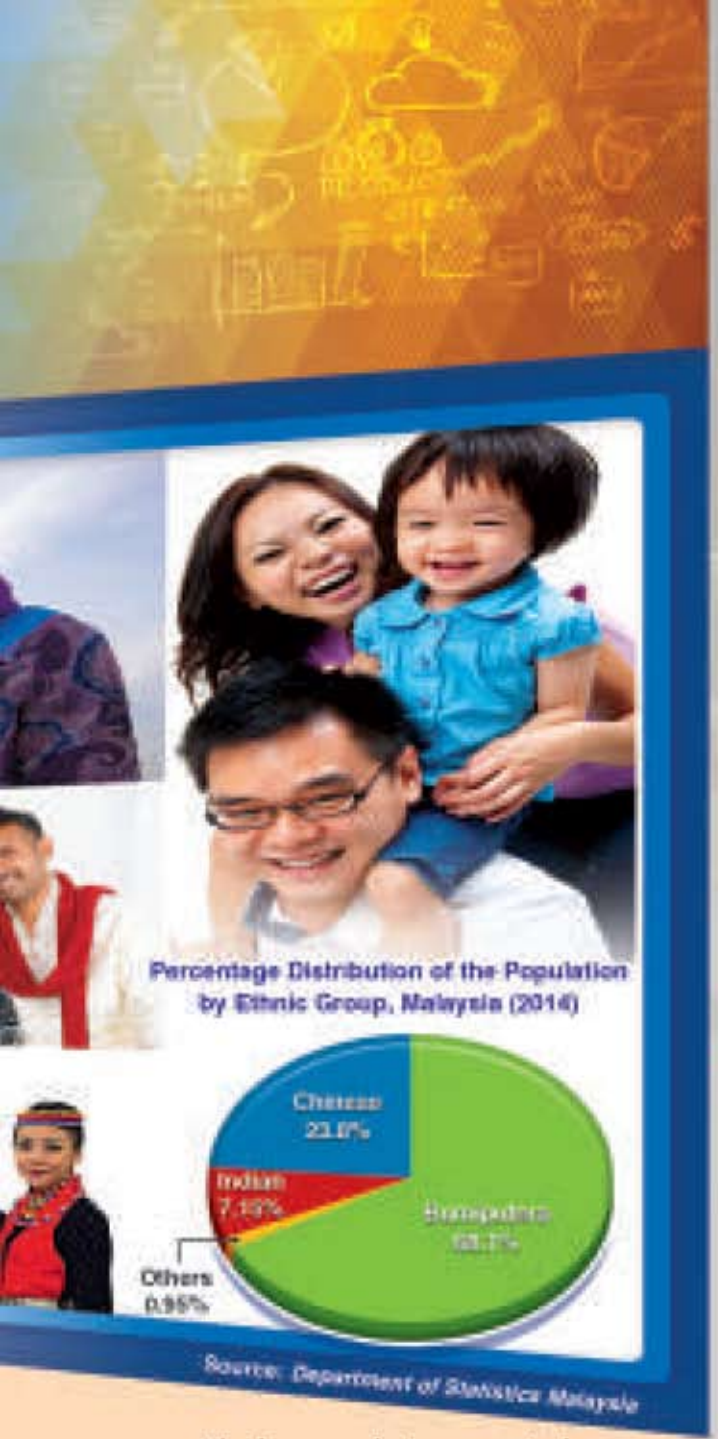
What will you learn?

- Data Collection, Organisation and Representation Process, and Interpretation of Data Representation

Why study this chapter?

As a basic knowledge in the field of statistics. Data handling involves skills like collecting, organising, representing, analysing and interpreting data, and then communicating the results of the data. Data handling is important as it enables us to understand the usage of data in newspapers, television and higher educational institutions as well as in the careers that we will pursue in the future. Discuss with your teacher other daily situations that involve data handling.

Malaysia is a multiracial country whereby all the people live together in peace and harmony. According to the Current Population Estimates 2014, the total population in Malaysia is 30.6 million; the number of males is 15.8 million and the number of females is 14.8 million.



The largest ethnic group is the *Bumiputera* who make up 68.1 percent of the total population, followed by Chinese (23.8%), Indian (7.15%) and others (0.95%).

Other than the data above, what other data can you obtain from the population distribution in Malaysia? How can this data be obtained?

Walking through Time


John Graunt

In ancient times, statistics were used by rulers to gather information about the population during their reign. However, statistics were only recorded in printed form in the 18th century by an English statistician, John Graunt. Two other English statisticians who actively contributed to the early development of statistics were Karl Pearson (1857 – 1936) and Ronald Fisher (1890 – 1962).

For more information:

http://goo.gl/Nx43ay

Word Link

• pie chart	• *carta pai*
• bar chart	• *carta palang*
• categorical data	• *data kategori*
• numerical data	• *data numerik*
• line graph	• *graf garis*
• histogram	• *histogram*
• frequency table	• *jadual kekerapan*
• displaying data	• *memaparkan data*
• analysing data	• *menganalisis data*
• classifying data	• *mengklasifikasikan data*
• organising data	• *mengorganisasikan data*
• collecting data	• *mengumpulkan data*
• interpreting data	• *mentafsir data*
• representing data	• *mewakilkan data*
• stem-and-leaf plot	• *plot batang-dan-daun*
• dot plot	• *plot titik*
• frequency polygon	• *poligon kekerapan*
• statistical question	• *soalan statistik*

Open the folder downloaded from page vii for the audio of Word Link.

12.1 Data Collection, Organisation and Representation Process, and Interpretation of Data Representation

How do you generate statistical questions and collect relevant data?

LEARNING STANDARDS

Generate statistical questions and collect relevant data.

To collect relevant data, we must generate statistical questions. What is a statistical question?

A **statistical question** is a question that can be answered by collecting data and there will be variability or diversity in the data related to the question. For example,

Statistical question	Reason
"How tall are the students in Class 1 Amanah?"	There is variability in the height of the students from that class, for example, 150 cm, 156 cm, 164 cm and so on.
"What is the favourite food of the students in Class 1 Amanah?"	There is variability in the types of food chosen by the students from that class, for example, *nasi lemak*, fried noodles, *laksa* and so on.

Information about the height of students and their favourite food are obtained through data collection.

Is each of the following questions a statistical question? Explain.

- *"How tall is Rosmee?"*
- *"Do the students in Class 1 Amanah like nasi lemak more than fried noodles?"*

Communication Corner

Generate statistical questions and collect data for each of the following in your class.

- Favourite sports of the students
- Modes of transport of students to school

After generating the statistical questions, the next step is to determine the method of data collection. Data can be collected by various methods.

DATA COLLECTION METHOD

Interview

For example, to find out how the students travel to school.

Survey

For example, to find out the favourite television programmes of the students.

Observation

For example, to count how many cars pass through the junction every hour.

Experiment

For example, to record the temperature of hot water as it cools down every five minutes.

Let's Discuss

Have a discussion with your friends about choosing the most suitable method to collect data. State the advantages and disadvantages of each method by giving examples to support your answers.

How do you classify data and construct frequency tables?

LEARNING STANDARDS

Classify data as categorical or numerical and construct frequency tables.

After collecting the data, the next step is classifying the data. Data can be classified into categorical data and numerical data.

DATA

Categorical	Numerical
• Measures characteristic • Cannot be measured numerically but can be described. For example, the gender of a person, colours of cars, flavours of sweets, blood groups	• Measures quantity • Measured numerically. For example, the number of books read in a week, the height of the badminton players, the time spent on exercise

SMART TIPS

Numerical data consists of
- **discrete data** that is measured in a whole unit. For example, the number of family members is 6 people.
- **continuous data** that is measured on a continuous scale. For example, the mass of the students are 53 kg, 56.2 kg and 66.5 kg.

Example 1

Classify the following data as categorical data or numerical data:
(a) The body temperature of each student
(b) The number of trees planted in each district
(c) The causes of road accidents

Solution

(a) Numerical data (b) Numerical data (c) Categorical data

Among the numerical data above, which is a discrete data and which is a continuous data?

SMART TIPS

Ungrouped data is an unprocessed raw data.

After classifying the data, the next step is to organise the ungrouped data by constructing a frequency table.

Example 2

The data shows the number of children in each family for 20 families. Organise the data by constructing a frequency table.

2	0	1	1	2	1	3	0	4	3
2	4	1	0	2	1	0	2	2	3

Solution

Number of children	Tally	Frequency
0	IIII	4
1	𝍸	5
2	𝍸 I	6
3	III	3
4	II	2
	Total	20

Self Practice 12.1a

1. Classify the following data as categorical data or numerical data:
(a) Number of stamps each student collects
(b) Time spent on the Internet
(c) Ability to play *sepak takraw*
(d) Colour of cars
(e) Length of earthworms
(f) Number of tourists visiting the National Museum each month
(g) Language spoken at home
(h) Annual income

2. The data shows the T-shirt size of Form 1 Cekal students. Organise the data by constructing a frequency table.

XL	L	XL	M	M	L	M	L	M
M	M	M	XL	XL	L	XL	L	M
M	L	M	L	L	S	M	M	L

How do you construct data representations?

LEARNING STANDARDS
Construct data representation for ungrouped data and justify the appropriateness of a data representation.

Data shown in the form of a table can also be presented graphically to make it easier to read and understand. The suitability of a data representation depends on the type of data collected and the purpose of acquiring the information. The data can be represented by a bar chart, a pie chart, a line graph, a dot plot and a stem-and-leaf plot.

(a) Bar chart

A bar chart is a type of data representation which represents data by using bars. It is suitable for showing comparisons between categories.

Example 3

The frequency table shows the activities during leisure time for Form 1 Bakti students. Construct a bar chart to represent the data and justify the appropriateness of the data representation.

Activity	Frequency
Reading	8
Watching television	9
Surfing the Internet	7
Exercising	6
Listening to music	4

SMART TIPS

The bars in a bar chart can be drawn horizontally or vertically.

Solution

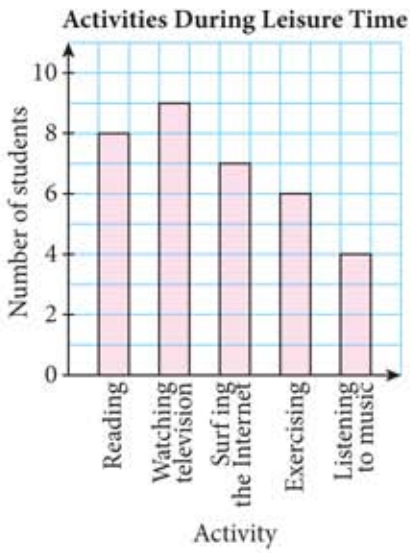
Steps to construct a bar chart:

Draw the horizontal and vertical axes on a grid.

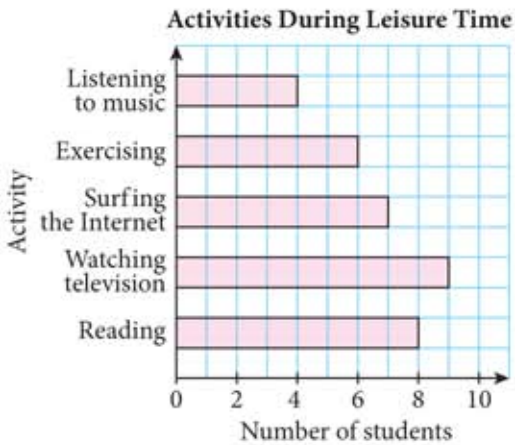
Choose one of the axes to mark a suitable scale and label the axis with the number of students. Label the other axis with types of activities.

Draw the bars such that the height of each bar corresponds to the frequency of the category it represents.

Write down the title of the bar chart.



or



These bar charts are suitable for comparing the number of students with the different leisure activities.

Scan the QR Code or visit **https://goo.gl/UxjF6c** and open the file *various bar charts_pdf* regarding data representation using various types of bar charts.

SMART TIPS

When representing data using a bar chart:

- the width of each bar must be uniform.
- the bars need to be evenly spaced.

Example 4

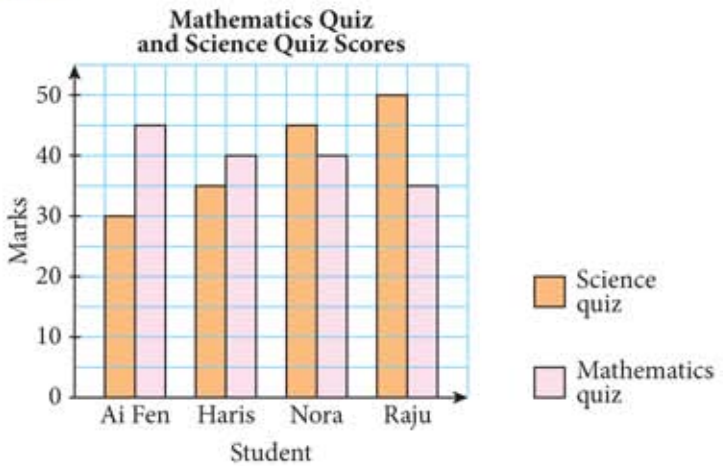
The table shows the scores obtained by a group of students in the Science quiz and the Mathematics quiz. Construct a bar chart to represent these two sets of data.

Student	Marks	
	Science quiz	Mathematics quiz
Ai Fen	30	45
Haris	35	40
Nora	45	40
Raju	50	35

Let's Discuss

Discuss whether a bar chart is suitable for representing data that has one or two categories that dominate the findings.

Solution



SMART TIPS

A dual bar chart is suitable for comparing two sets of data. For example, the performance of students in two tests, the price of hotel rooms at normal rate and during school holidays.

(b) Pie chart

A pie chart is a data representation that uses sectors of a circle to show the portion of each category of the whole data.

Example 5

The table shows the number of cars of Dynamic model sold by a car dealer. Construct a pie chart to represent the data and justify the appropriateness of the data representation.

Colour of cars	Red	Yellow	White	Blue
Number of cars	9	12	10	5

Let's Discuss

Discuss whether a pie chart is suitable for representing data that has many categories or the fraction of each category is fairly equal.

Solution

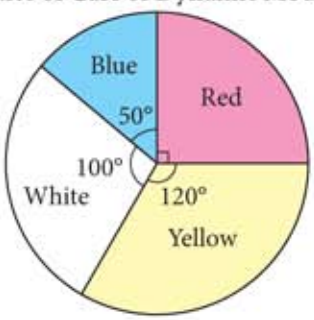
Steps to construct a pie chart:

Find the angle of sector for each category. → Draw a circle and divide it into different sectors based on the angles calculated. → Label each sector. → Write down the title of the pie chart.

Colour of cars	Number of cars	Fraction of the circle	Angle of the sector
Red	9	$\frac{9}{36}$	$\frac{9}{36} \times 360° = 90°$
Yellow	12	$\frac{12}{36}$	$\frac{12}{36} \times 360° = 120°$
White	10	$\frac{10}{36}$	$\frac{10}{36} \times 360° = 100°$
Blue	5	$\frac{5}{36}$	$\frac{5}{36} \times 360° = 50°$
Total	36	1	360°

Angle of the sector $= \frac{\text{Frequency of data}}{\text{Total frequency}} \times 360°$

Sales of Cars of Dynamic Model



This pie chart is suitable for comparing each colour of the cars with the total number of cars.

(c) Line graph

A line graph is a data representation used to display changes of data over a period of time. The data is represented by points which are connected in a straight line.

Example 6

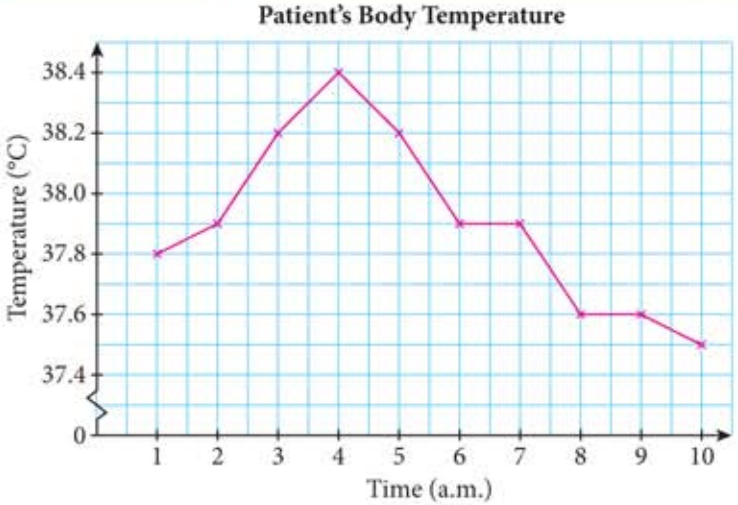
The table shows the temperature of a patient over a particular period of time. Construct a line graph to represent the data and justify the appropriateness of the data representation.

Time (a.m.)	1	2	3	4	5	6	7	8	9	10
Temperature (°C)	37.8	37.9	38.2	38.4	38.2	37.9	37.9	37.6	37.6	37.5

Solution

Steps to construct a line graph:

Draw the horizontal and vertical axes on a grid. → Choose a suitable and uniform scale for both axes. The vertical axis represents data. The horizontal axis represents time. → Plot the points and connect the points in a straight line. → Write down the title of the line graph.



This line graph is suitable for displaying changes in a patient's body temperature over a period of 10 hours.

SMART TIPS

For a line graph, the horizontal axis usually represents the time duration and the vertical axis usually represents the frequency values.

Let's Discuss

Discuss whether a line graph is suitable for
- predicting data trends.
- showing clearly the fluctuations of the data before and after a certain point of time.

(d) Dot plot

A dot plot shows the distribution of data on a number line. The data are either clustered around certain values or spread out evenly on a number line. Dot plot can help us to visualise data patterns, draw inferences and make decisions. Dot plot can also detect unusual observations, that is, the extreme values in the data. If there is extreme value in the data, we may need to investigate further to find out the cause for the unusual observations.

Career in Mathematics

A basketball coach uses a dot plot to evaluate the performance of each player under his supervision.

Example 7

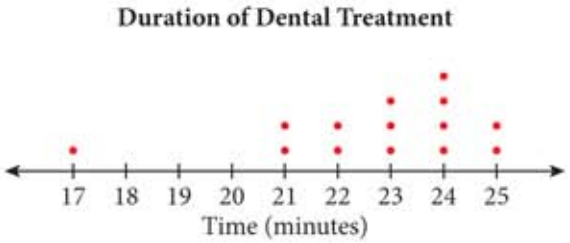
The duration of dental treatment (in minutes) provided by a dentist to 14 patients is as shown below. Represent the data in a dot plot and justify the appropriateness of the data representation.

23	24	21	24	25	24	25
24	22	17	21	23	22	23

Solution

Steps to construct a dot plot:

Draw a horizontal number line that covers the range of the given data. → Plot the individual data with a dot over their values on the number line. → Write down the title of the dot plot.



This dot plot is suitable for displaying the duration of the dental treatment which ranged from 17 minutes to 25 minutes.

Let's Discuss

Discuss whether a dot plot is suitable for
- illustrating the data frequency.
- displaying a large amount of data.
- representing categorical data or numerical data.

(e) Stem-and-leaf plot

A stem-and-leaf plot is a data representation that separates the data values into stem and leaf according to their place value. The leaf usually is the last digit of the number. The stem is the remaining digit or digits on the left of the number. The plot retains the original data values. Therefore, we are able to do arithmetic calculation on these values for the purpose of data analysis.

Walking through Time

The American mathematician John W. Tukey (1915-2000) introduced the stem-and-leaf plot in the 1960s. Since then, the stem-and-leaf plot has become a popular way of representing data for analysis.

Example 8

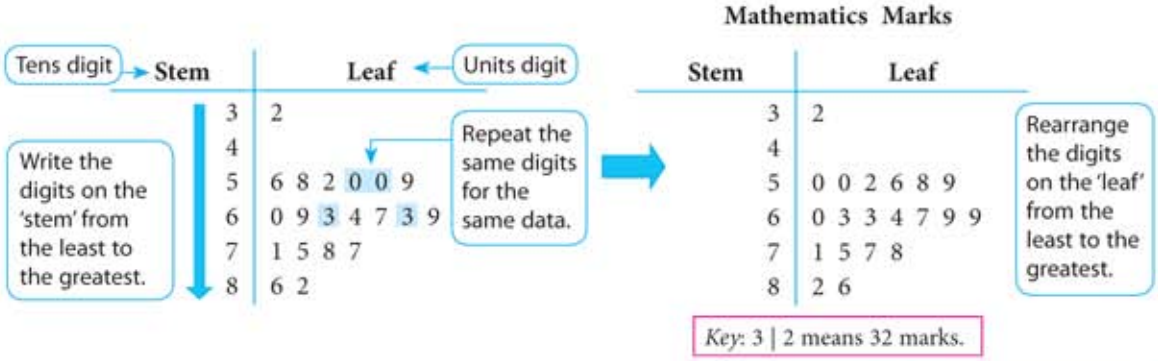
The data shows the Mathematics marks for a class of 20 students. Represent the data by using a stem-and-leaf plot and justify the appropriateness of the data representation.

60	56	69	32	63	58	71	86	52	64
50	67	82	63	75	50	69	78	77	59

Solution

Steps to construct a stem-and-leaf plot:

Write each data one by one and take the tens digit of each data value as the stem. → The last digit of the number is written on the leaf. → Rearrange the leaves in ascending order. → Write a key and the title. The key indicates the unit for the stem and leaf.



This stem-and-leaf plot is suitable for displaying the marks of each student in the class.

Smart Technology

Open the folder downloaded from page vii to open the file *Bar Chart.xls* for the data representation in Example 3. Hence, continue your exploration with other types of data representation.

Let's Discuss

Discuss whether the stem-and-leaf plot is suitable for

- displaying large data
- representing categorical data or numerical data

Self Practice 12.1b

1. A survey was carried out in a class to find out how the students travel to school, and the results are as shown in the table below. Construct a bar chart to represent the data and justify the appropriateness of the data representation.

Transportation	Car	School bus	Public bus	Bicycle	Walk
Frequency	8	10	7	2	5

2. The table shows the prices of four types of accommodation around the Historical City of Melaka at normal rate and during holiday season. Construct a bar chart to represent the two sets of data and justify the appropriateness of the data representation.

Type of accommodation	Hotel	Homestay	Budget hotel	Hostel
Normal rate (RM)	300	250	150	100
Holiday season (RM)	350	300	200	100

3. The table shows the favourite songs of a group of children.

Favourite song	Rasa Sayang	Ikan Kekek	Bangau Oh Bangau	Geylang Si Paku Geylang	Lompat Si Katak Lompat	Dayung Sampan
Number of children	30	40	20	15	10	5

Construct a pie chart to represent the data and justify the appropriateness of the data representation.

4. The table shows Kamil's height over a period of six years. Construct a line graph to represent the data and justify the appropriateness of the data representation.

Year	2011	2012	2013	2014	2015	2016
Height (cm)	145	150	153	160	164	167

5. The data shows the number of text messages sent by a group of students on a particular day. Represent the data by using a dot plot and justify the appropriateness of the data representation.

3	4	8	7	11
6	5	7	6	3
9	6	5	11	8

6. In a survey, the ages of 24 readers of a magazine are recorded as follows. Represent the data by using a stem-and-leaf plot and justify the appropriateness of the data representation.

44	53	33	65	51	30	42	34
57	36	51	32	39	44	25	31
58	47	31	22	58	38	60	47

How do you convert a data representation to another representation?

LEARNING STANDARDS

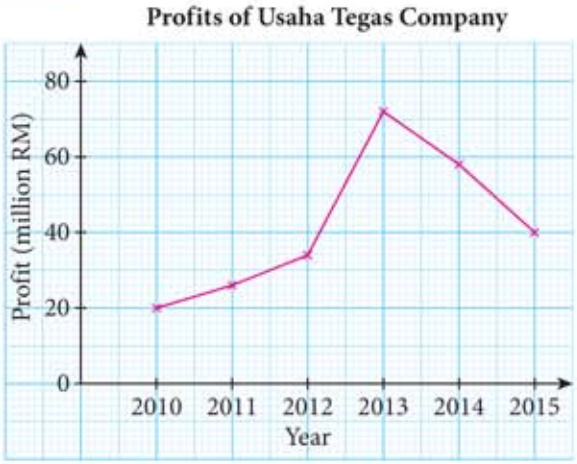
Convert a data representation to other suitable data representations with justification.

A data representation can be converted to other suitable representations for further analysis.

Example 9

The bar chart shows the profits obtained by Usaha Tegas Company from 2010 to 2015. Convert the bar chart representation to another suitable representation and justify the conversion for this representation.

Solution



The bar chart representation is converted to a line graph representation because the line graph is suitable for displaying data collected over a certain period, that is, the profit performance of Usaha Tegas Company over a period of six years.

Example 10

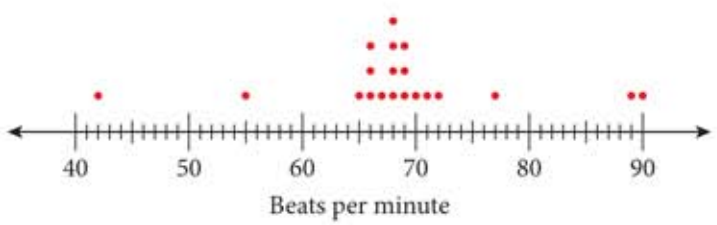
The stem-and-leaf plot shows the results of a survey on the pulse rates per minute of patients treated at a community polyclinic. Convert the representation to a dot plot and justify the conversion for this representation.

Pulse Rates Per Minute of Patients

Stem	Leaf
4	2
5	5
6	5 6 6 6 7 8 8 8 8 9 9 9
7	0 1 2 7
8	9
9	0

Key: 4 | 2 means 42 beats per minute.

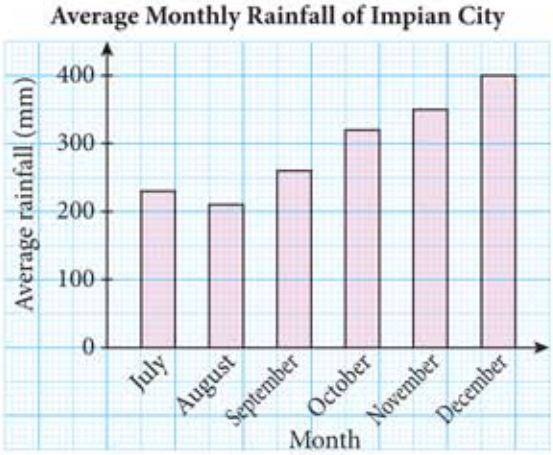
Solution



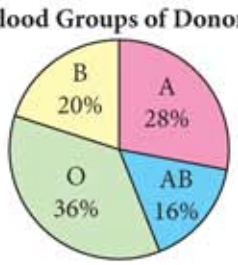
The stem-and-leaf plot is converted to a dot plot because both of these representations are suitable for displaying numerical data distribution and at the same time retain the original data values.

Self Practice 12.1c

1. The bar chart shows the average monthly rainfall of Impian City from July to December 2015. Convert the representation to another suitable data representation and justify the conversion.



2. A blood donation campaign organised by the Malaysian Red Crescent has received an overwhelming response from the public. The pie chart shows the blood groups of 25 donors in the first three hours. Convert this representation to another suitable data representation and justify the conversion.



How do you interpret data representations?

By interpreting data representations, we can obtain information and hence make inferences and predictions.

LEARNING STANDARDS

Interpret various data representations including making inferences or predictions.

Example 11

The line graph shows the mass of waste, in thousand tonnes, produced in a city from 2010 to 2015.

(a) What is the mass of waste produced in 2010?

(b) What can you say about the mass of waste produced in 2011 and 2014?

(c) Find the mean mass of the waste produced over a period of six years.

(d) State one inference based on the line graph given.

(e) Based on the trend of the line graph, predict the mass of waste produced in 2016.

Solution

(a) 1 900 tonnes

(b) The mass of waste produced in 2011 and 2014 are the same.

(c) Total mass of waste produced over a period of 6 years
$= 1\,900 + 2\,100 + 2\,900 + 2\,600 + 2\,100 + 1\,600$
$= 13\,200$ tonnes

$$\text{Mean mass of waste} = \frac{13\,200}{6}$$
$$= 2\,200 \text{ tonnes}$$

(d) The mass of waste produced decreases each year after the year 2012.

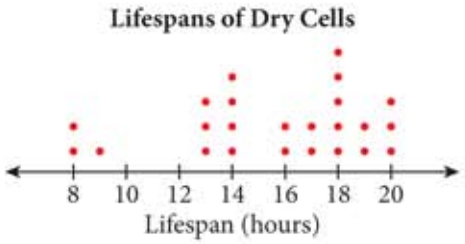
(e) 1 100 tonnes

Communication Corner

Think and state the possible reasons for the gradual decrease in waste production in a city year by year.

Example 12

In a quality control laboratory, the lifespans (to the nearest hours) of 24 dry cells are tested. The data obtained is represented by a dot plot as shown.



(a) State the maximum and minimum lifespans of the dry cells tested.

(b) State one inference based on the data from the dot plot.

(c) The quality control laboratory has decided that the dry cells with a lifespan of less than 10 hours will be considered defective and thus will be rejected. Find the percentage of the dry cells that will be rejected.

(d) It is known that 50% of the dry cells have a lifespan of at least x hours. Find the value of x.

Solution

(a) The maximum lifespan = 20 hours
The minimum lifespan = 8 hours

(b) Most of the dry cells have a lifespan from 13 hours to 20 hours.

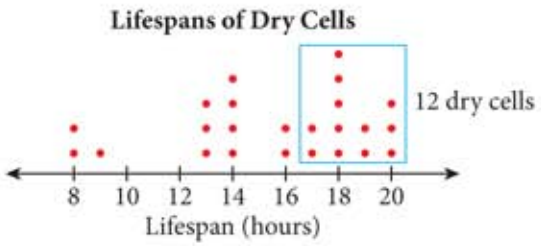
(c) Number of dry cells with a lifespan of less than 10 hours = 3

$$\text{Percentage of dry cells rejected} = \frac{3}{24} \times 100\%$$
$$= 12.5\%$$

(d) 50% of the number of dry cells

$$= \frac{50}{100} \times 24$$
$= 12$ dry cells

From the dot plot, there are 12 dry cells with a lifespan of at least 17 hours.
Thus, $x = 17$



Example 14

The frequency polygon shows the time taken by a group of participants to complete the fun run in a park.

(a) Find the total number of participants in the fun run.

(b) Find the number of participants who recorded a time of 16 minutes to 20 minutes to complete the run.

(c) By observing the shape of the frequency polygon, make an inference based on the distribution of time taken by the participants.

Solution

(a) Total number of participants = 10 + 24 + 30 + 26 + 12 + 8
= 110 participants

(b) 24 participants ← 24 participants recorded a time of 16 – 20 minutes

(c) Most of the participants recorded a time of 16 minutes to 30 minutes to complete the run.

Self Practice 12.1d

1. A botanist studies the height of a tree in a tropical rainforest. The line graph shows the height of the tree over a period of seven years.

(a) How tall was this tree initially?

(b) What is the increase in height of the tree over a period of seven years?

(c) In which year will the height of the tree be 12 m?

(d) State an inference based on the line graph given.

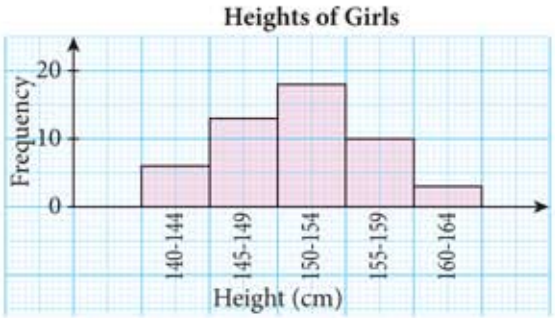
(e) Based on the trend of the line graph, predict the height of the tree in 2017.

Histogram

A histogram is a data representation that displays grouped data. Grouped data is data that is collected in intervals.

Example 13

The histogram shows the heights of 50 girls.



(a) Find the number of girls with heights from 155 cm to 159 cm.

(b) A girl with a height of 160 cm and above is eligible to participate in the volleyball team. Find the number of girls who are eligible to participate in the volleyball team.

(c) By observing the shape of the histogram, make an inference based on the distribution of the heights of the girls.

Solution

(a) 10 girls

(b) 3 girls ← There are 3 girls with heights of 160 – 164 cm.

(c) Most of the girls have heights of 145 cm to 159 cm.

Did You Know

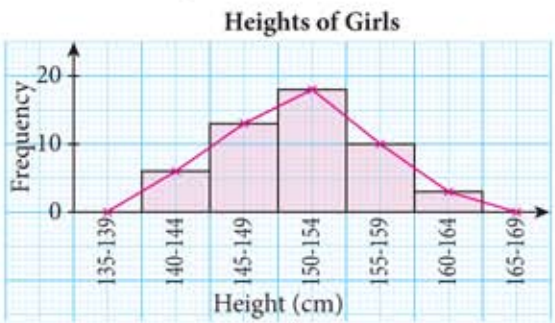
- The width of each bar in a histogram represents a specific interval. For example, the interval 140-144 covers the range of heights from 140 cm to 144 cm.
- The height of each bar represents the frequency in each interval.

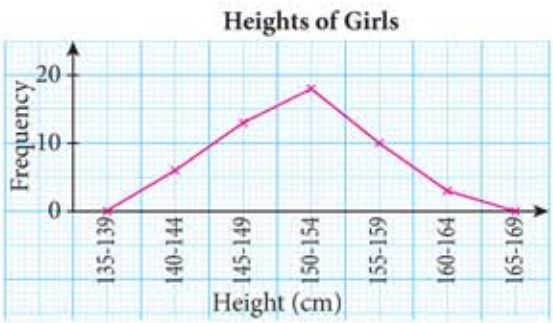
SMART TIPS

- A histogram does not display the actual values of the data but displays values in a certain interval.
- A histogram can provide a display of large data sets because the data is represented in class intervals.

Frequency polygon

A frequency polygon is a graph formed by joining the midpoint of the top of each bar in a histogram with straight lines. Based on Example 13, a frequency polygon can be drawn from the histogram and is shown below.





2. The stem-and-leaf plot shows the diameters of axles manufactured by a machine.

Diameters of Axles

Stem	Leaf
24	5 6 6 8
25	0 1 1 2 3 6
26	0 0 0 1 3 3 4 5 8 9
27	3 4 4 5 7 8 8
28	2 3 6

Key: 24 | 5 means 24.5 mm.

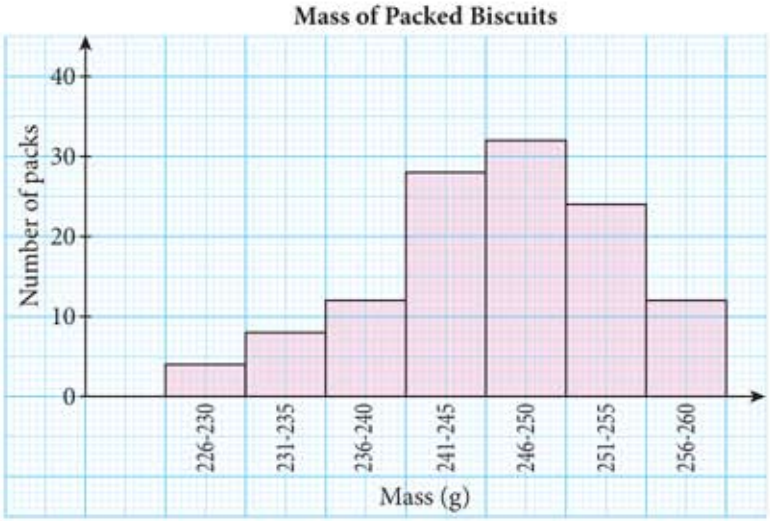
(a) Determine the total number of axles manufactured.

(b) Find the largest and the smallest diameters of the axles manufactured.

(c) A mechanic notices that he has to reduce the axles with diameters exceeding 27.5 mm so that the axles can fit into wheels. Calculate the percentage of axles that need to have their diameters reduced.

(d) State an inference based on the distribution of the diameters of axles in this stem-and-leaf plot.

3.



A quality control supervisor wants to determine whether the batch of biscuits produced by the packaging department follows the specifications of the standard mass. The histogram shows the mass of packed biscuits in a few sample tests.

(a) How many packs of biscuits are tested in the sample?

(b) How many packs of biscuits have a mass from 236 g to 240 g?

(c) Based on the specifications, if 75% of the sample have mass from 241 g to 260 g, then the batch of biscuits produced would be packaged and sent for shipment. Does this batch of biscuits produced satisfy the specifications? Show your calculations.

What is the importance of representing data ethically?

LEARNING STANDARDS

Discuss the importance of representing data ethically in order to avoid confusion.

Data representations help us to analyse and interpret data much easier. We need to represent the data ethically to avoid confusion.

To represent the data ethically,

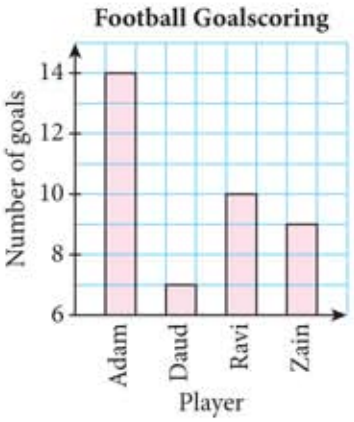
- the scale used in the representation must be consistent and start at 0.
- the data displayed must be accurate.

Example 15

The bar chart shows the number of goals scored by four football players in the Permai District Football League.

(a) Is the number of goals scored by Adam twice the number of goals scored by Ravi? Explain your answer.

(b) In your opinion, does this bar chart clearly represent the number of goals scored by the players?



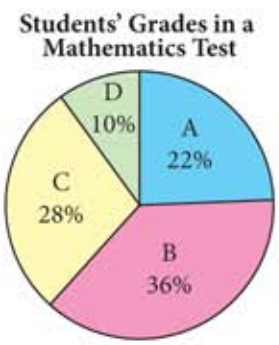
Solution

(a) Number of goals scored by Adam = 14
Number of goals scored by Ravi = 10
Thus, the number of goals scored by Adam is not twice the number of goals scored by Ravi.

(b) No, the information displayed is misleading. The scale in the vertical axis should start from 0.

Example 16

The pie chart shows the grades obtained by a group of students in a Mathematics test. Does this pie chart display data accurately? Explain your answer.



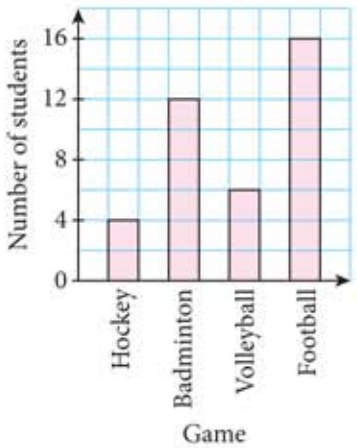
Solution

Total percentage = 22 + 36 + 28 + 10
= 96

No, because the total percentage in the pie chart is not equal to 100.

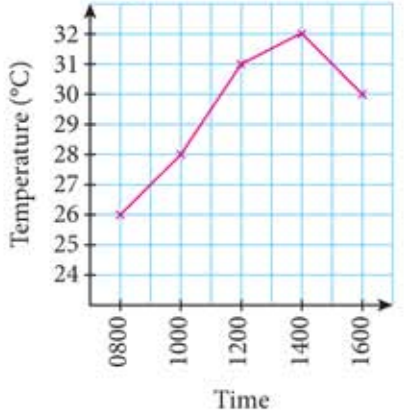
Self Practice 12.1e

1.



Zurini conducts a survey on 40 students to find out their favourite games. The results of the survey are shown in the bar chart. Does this bar chart display data accurately? Explain your answer.

2.

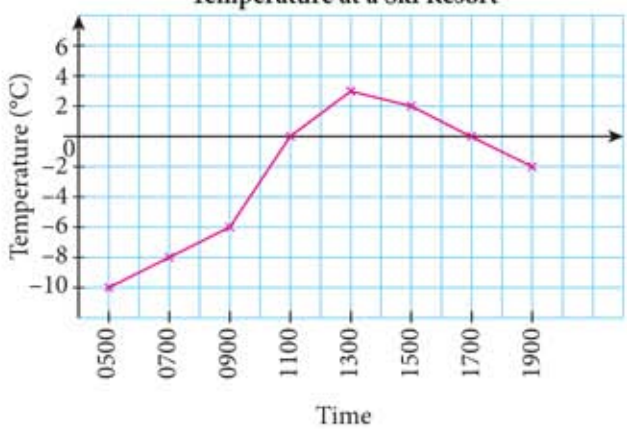


The line graph shows the temperatures in Mahkota City from 0800 to 1600. Does this line graph clearly display the data? Explain your answer.

Mastery Q 12.1

Open the folder downloaded from page vii for extra questions of Mastery Q 12.1.

1.



The line graph shows the temperatures recorded at a ski resort on a particular day during the ski season.

(a) Estimate the temperature at 1000.
(b) Predict the temperature at 2100.
(c) At what time was the temperature 0 °C?

2.

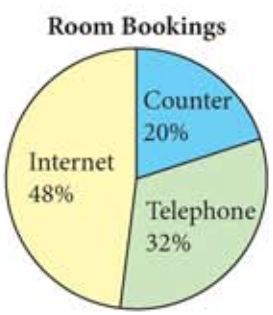
Animal	Cheetah	Zebra	Lion	Horse	Deer	Ostrich
Speed (km/h)	110	65	80	75	70	95

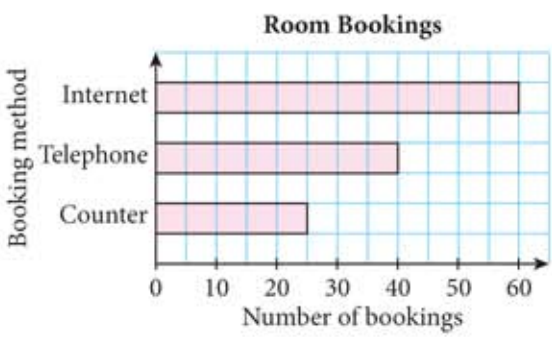
The frequency table shows the maximum speed, in km/h, of several animals.

(a) Represent the above data in a
  (i) bar chart,
  (ii) stem-and-leaf plot.

(b) Between the two types of data representations constructed in (a), which representation is more suitable? Explain your answer.

3.





A hotel has received 125 room bookings via three methods on a certain day. The data is displayed by using a pie chart and a bar chart.

(a) What is the difference between the number of room bookings via the Internet and the number of room bookings over the counter? Which representation shows this data more clearly? Explain your answer.

(b) Almost half of the room bookings are made via the Internet. Which representation shows this data more clearly? Explain your answer.

(c) Which representation clearly shows the number of room bookings made via the Internet? Justify your answer.

(d) Do you think a line graph is suitable for displaying this data? Explain your answer.

(e) What other representation is suitable for displaying this data?

4. The data shows the daily allowances (in RM) of a group of factory workers.

20	25	21	24	22	23
22	22	23	30	25	22

(a) Represent the data in a dot plot.

(b) Describe briefly
  (i) the data distribution for the daily allowances of the group of factory workers,
  (ii) the value where most of the data are clustered around,
  (iii) whether there is any extreme value in the data.

SUMMARY

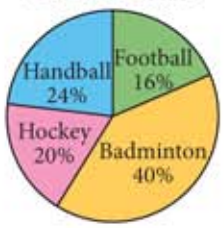
DATA HANDLING

- **Data**
  - Categorical
  - Numerical
- **Frequency**
  - Frequency table

DATA REPRESENTATION

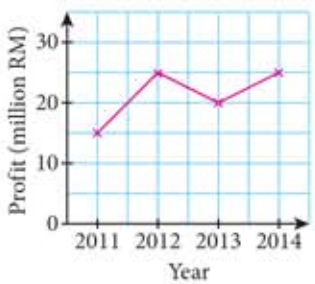
Pie chart

Favourite Game



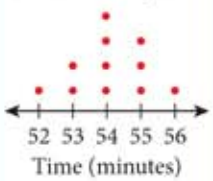
Line graph

Profits of Maju Company



Dot plot

Training Time for a Cross-country Run



Stem-and-leaf plot

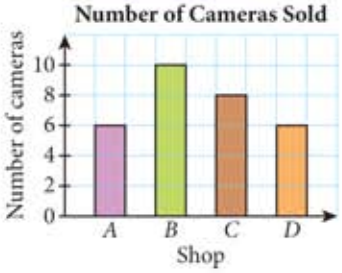
Basketball Points

Stem	Leaf
0	2
1	2 2 2 5 5 8
2	0 0 1 1 1 4 6 6 7 8
3	0 0

Key: 1 | 5 means 15 points

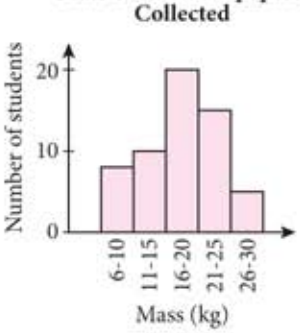
Bar chart

Number of Cameras Sold



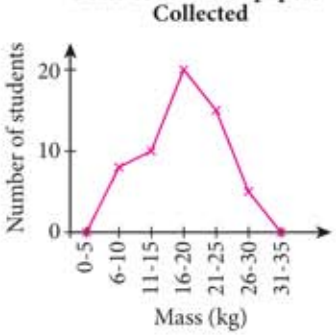
Histogram

Mass of Old Newspapers Collected



Frequency polygon

Mass of Old Newspapers Collected



5. Diagram (a) and Diagram (b) show two histograms which represent the Science and Mathematics marks of Form 1 students.

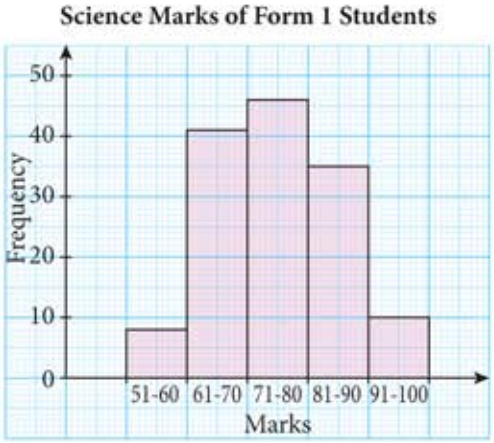


Diagram (a)

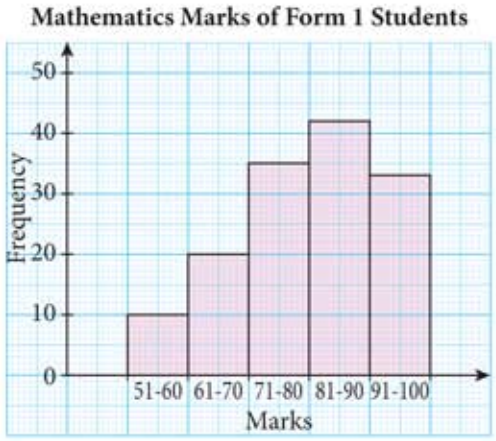


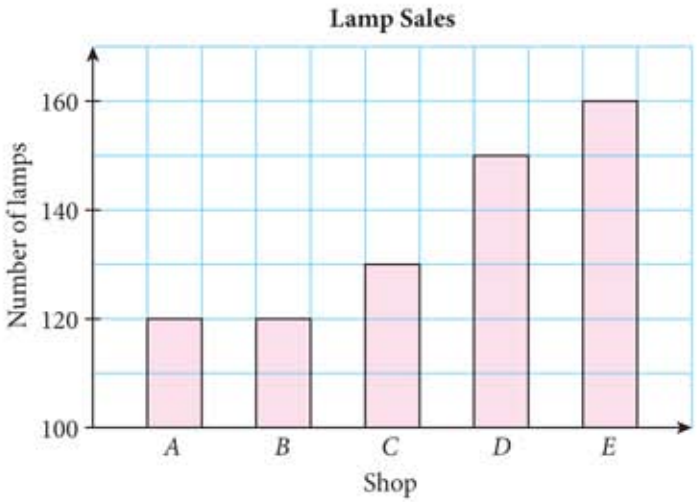
Diagram (b)

(a) Find the number of students who obtained more than 80 marks in each test.

(b) Describe briefly the distribution of marks in each test.

(c) Compare the distributions of the Science and Mathematics marks. What inference can be made?

6. The bar chart shows the number of lamps sold by five shops in a month.



(a) What can you say about the number of lamps sold by shop *A* and shop *B*?

(b) Shop *E* claims that the number of lamps it sold is twice the number of lamps sold by Shop *C*. Is this claim valid? Explain your answer.

At the end of this chapter, I can...	Very good	Work harder
generate statistical questions and collect relevant data.		
classify data as categorical or numerical and construct frequency tables.		
construct data representation for ungrouped data and justify the appropriateness of a data representation.		
convert a data representation to other suitable data representations with justification.		
interpret various data representations including making inferences or predictions.		
discuss the importance of representing data ethically in order to avoid confusion.		

Let's PRACTISE

Test Yourself

1. The line graph shows the initial selling prices of house A and house B and the rate of increase in prices over a 10-year period.

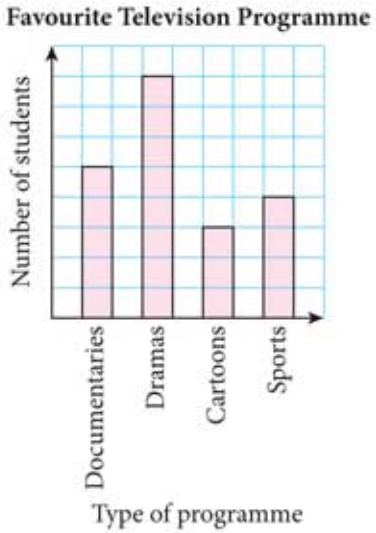
(a) Which house has a higher rate of increase in price?

(b) In which year will the prices of house A and house B be equal?

2. The bar chart shows four types of favourite television programmes of a group of students.

(a) State the most favourite television programme of the group of students.

(b) Represent all the information on the bar chart by using a pie chart.

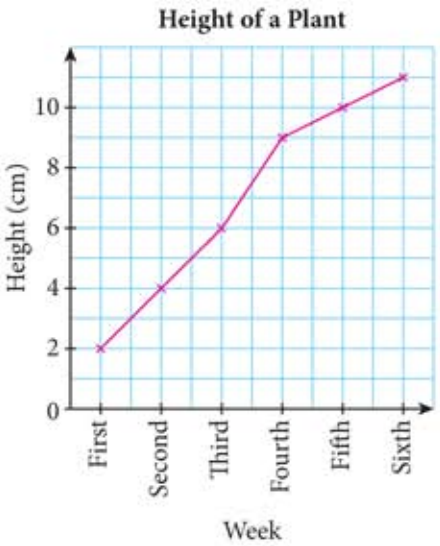


3. In an experiment, Johari measures the height of a plant every week over a six-week period. He then presents the data collected in a line graph.

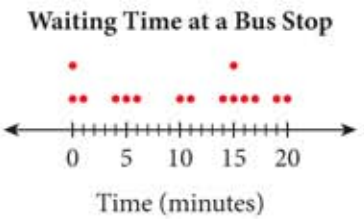
(a) State which weeks the plant will grow at the same rate.

(b) Between which two weeks will the plant attain the largest increase in height?

(c) What is the advantage of this representation?



4. A survey was conducted to study the waiting time (to the nearest minutes) of a group of passengers at a bus stop. The data obtained from the survey is represented in a dot plot. Based on the dot plot, determine whether the following statements are TRUE or FALSE:



(a) 20 passengers were involved in this survey.

(b) The shortest waiting time was 0 minute.

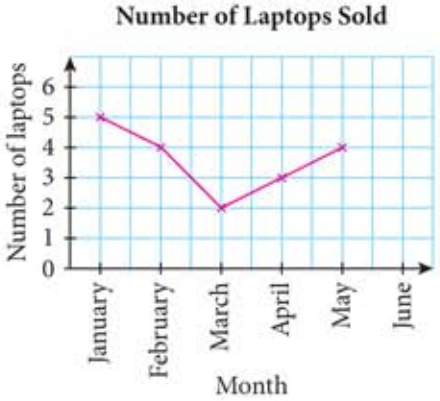
(c) 40% of the passengers faced a waiting time of at least 15 minutes.

Self Mastery

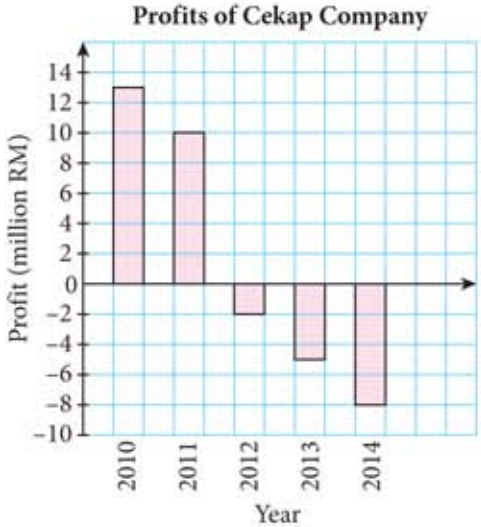
5. The incomplete line graph shows the sales of laptops in a computer shop from January to May. The number of laptops sold in June is not shown.

(a) The total number of laptops sold from January to June is 24 units. Copy and complete the line graph for June.

(b) Convert the line graph to another suitable representation and justify the conversion of this representation.



6. The bar chart shows the profits of Cekap Company from 2010 to 2014.



(a) In which year did Cekap Company obtain maximum profit? What was the maximum profit?

(b) In which year did Cekap Company start incurring losses? What was the loss in that year?

(c) What was the profit or loss in 2014?

(d) (i) In 2015, given that Cekap Company recorded an increase of RM11 million in profit compared to 2014. Based on the information, complete the bar chart for 2015 on the same diagram.

(ii) Hence, calculate the company's total profits or losses over a six-year period.

Challenge Yourself

7. The time taken (in minutes) by a plumber to fix 30 leaking pipes are shown in the stem-and-leaf plot below.

Time Taken to Fix Leaking Pipes

Stem	Leaf
1	2 5 5 6 7 8
2	1 4 5 7 8
3	0 0 1 3 4 5 6 7 7 8 9
4	1 2 3 5 7 8
5	2 6

Key: 1 | 2 means 12 minutes.

(a) List all the data displayed in the stem-and-leaf plot above.

(b) State the shortest time taken by the plumber to fix a leaking pipe.

(c) What inference can you make regarding the times taken to fix the leaking pipes?

8.

Ticket Price of Desa Mutiara Theme Park

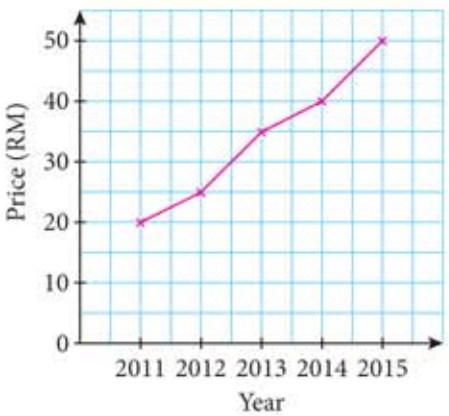


Diagram (a)

Ticket Price of Desa Mutiara Theme Park

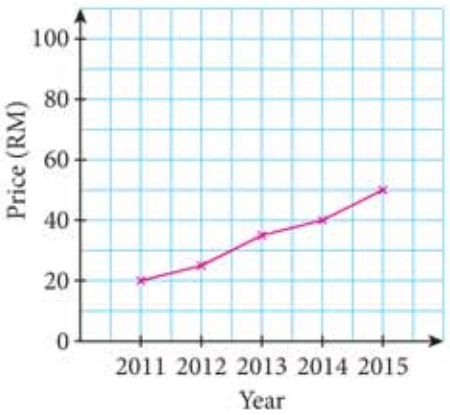


Diagram (b)

The line graphs in Diagram (a) and Diagram (b) display the same data for the ticket prices of Desa Mutiara theme park from 2011 to 2015.

(a) Which line graph shows a higher increase in price over a five-year period?

(b) Which line graph might the theme park manager use to show that the price increase is not significant? Is this an ethical representation? Explain your answers.

CHAPTER 12

ASSIGNMENT

In this assignment, you are to collect, display, analyse and interpret data regarding the media channels used to promote a new product among teenagers. The media channels include newspapers, television, radio, Internet, social media, magazines, catalogues, pamphlets and others.

Write a report to suggest the media channels that can extensively promote the new product among teenagers, and predict the media channels that would become increasingly popular in the future. To support your suggestion, your report must include questionnaires, frequency tables, suitable data representations using software, data interpretations and conclusions.

Exploring MATHEMATICS

Read the following article and discuss the questions posed.

PUTRAJAYA: The National Consumer Complaints Centre (NCCC) revealed that in the year 2012, they had received the highest number of complaints from the telecommunications category with 5 985 complaints. The second highest source of complaints was from the general consumer products category with 5 764 complaints followed by the utility-based category with 5 568 complaints.

No	Category of Complaints	Number of complaints
1	Telecommunications	5 985
2	General Consumer Products	5 764
3	Utility-based	5 568
4	Banking	5 555
5	Automobile	2 986
6	Travel and Leisure	2 034
7	Health and Fitness	1 839
8	Housing and Property	1 734
9	Vehicles Inspection – Workshops	1 695
10	Consumer Services – Retail	1 617

Source: Generasi Pengguna, August 2013, jointly published by the Ministry of Domestic Trade, Co-operatives and Consumerism, Malaysia and FOMCA

1. State the three categories that received the most number of complaints.
2. A statistician commented that "there is a possibility that the general consumer products category received more complaints than the telecommunications category". Discuss why this comment may be true.
   Note: You can think of how the data (number of complaints) have been arranged.
3. What can you tell about this data organisation?