Module 5: Data Handling

Unit 1  Data Collection

1.1 Types of Data

The objective of this section is to

• understand the differences between qualitative data and quantitative data.

**Qualitative data** is data that is not given numerically:

e.g. favourite colour, place of birth, favourite food, type of car.

**Quantitative data** is numerical. There are two types of quantitative data.

Discrete quantitative data can only take specific numeric values;

e.g. shoe size, number of brothers, number of cars in a car park.

Continuous quantitative data can take any numerical value;

e.g. height, mass, length.

Example

The following chart gives information about the two finalists in the 2013 mens' Wimbledon tennis championship.

Read through the information and answer these questions.

(a) Choose which of these terms

qualitative data

continuous quantitative data

discrete quantitative data

best describes the following information.

(i)  Age

(ii) Birthplace

(iii) Height

(iv) World Ranking

(v)  Aces

(vi) Fastest Serve

(vii) Love Life
(b) Find another attribute that can be described as

(i) qualitative data
(ii) continuous quantitative data
(iii) discrete quantitative data

<table>
<thead>
<tr>
<th>Novak Djokovic</th>
<th>2013</th>
<th>Andy Murray</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SRB)</td>
<td></td>
<td>(GBR)</td>
</tr>
<tr>
<td>26 years</td>
<td>Age</td>
<td>26 years</td>
</tr>
<tr>
<td>Belgrade, Serbia</td>
<td>Birthplace</td>
<td>Dunblane, Scotland</td>
</tr>
<tr>
<td>Monte Carlo, Monaco</td>
<td>Residence</td>
<td>London, England</td>
</tr>
<tr>
<td>Ivan Lendl</td>
<td>Coach</td>
<td>Marián Vajda</td>
</tr>
<tr>
<td>6 ft 2 in (1.88 m)</td>
<td>Height</td>
<td>6 ft 3 in (1.91 m)</td>
</tr>
<tr>
<td>176 lb (80 kg)</td>
<td>Weight</td>
<td>185 lb (84.1 kg)</td>
</tr>
<tr>
<td>1</td>
<td>World Ranking</td>
<td>3</td>
</tr>
</tbody>
</table>

**Wimbledon Final 2013**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aces</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Double Faults</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>65% First Serve Percentage</td>
<td>64%</td>
<td>59% First Serve Points Won</td>
</tr>
<tr>
<td>60% Second Serve Points Won</td>
<td>52%</td>
<td>127 mph Fastest Serve</td>
</tr>
<tr>
<td>118 mph First Serve Average</td>
<td>117 mph</td>
<td></td>
</tr>
<tr>
<td>97 mph Second Serve Average</td>
<td>180 mph</td>
<td>40 Unforced Errors</td>
</tr>
<tr>
<td>96 Total Points Won</td>
<td>114</td>
<td></td>
</tr>
</tbody>
</table>

**Lifestyle**

<table>
<thead>
<tr>
<th></th>
<th>Car: German Audi</th>
<th>Car: VW Polo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Love life:</td>
<td>His girlfriend, since 2005, is Jelena Ristic, director of a charity foundation</td>
<td>Is engaged to Kim Sears, a pet portraitist, his girlfriend since 2007</td>
</tr>
<tr>
<td>Likes:</td>
<td>Playing golf, skiing, working for charities</td>
<td>Football, does work for charities</td>
</tr>
</tbody>
</table>

**Solution**

(a) (i) Discrete quantitative, because it is given as a whole number.
(ii) Qualitative.
(iii) Continuous quantitative - it can take any value, although it is given here to the nearest inch (cm).
(iv) Discrete quantitative - it can only take positive whole numbers.
(v) Discrete quantitative.
(vi) Continuous quantitative - although it should be noted that it is given here as a whole number.

(vii) Qualitative - definitely!

(b) (i) Coach (ii) Weight (iii) Double faults

Exercises

1. Mr. Jenkin starts to make a database for his tutor group.

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Primary School</th>
<th>Transport to School</th>
<th>Height</th>
<th>Glasses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice Ascott</td>
<td>11</td>
<td>St. Johns</td>
<td>Bus</td>
<td>145 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Ben Bray</td>
<td>12</td>
<td>At. Andrews</td>
<td>Walk</td>
<td>160 cm</td>
<td>No</td>
</tr>
<tr>
<td>Carol Cotton</td>
<td>12</td>
<td>Prince Hill</td>
<td>Car</td>
<td>161 cm</td>
<td>No</td>
</tr>
<tr>
<td>David Darby</td>
<td>12</td>
<td>Prince Hill</td>
<td></td>
<td>152 cm</td>
<td>No</td>
</tr>
<tr>
<td>Eddie English</td>
<td>11</td>
<td>St. Andrews</td>
<td>Walk</td>
<td>158 cm</td>
<td>Yes</td>
</tr>
<tr>
<td>Frederick Franks</td>
<td>12</td>
<td>St. Andrews</td>
<td>Bike</td>
<td>164 cm</td>
<td>No</td>
</tr>
<tr>
<td>Graham Gee</td>
<td>12</td>
<td>St. Johns</td>
<td>Bus</td>
<td>166 cm</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(a) What is missing from Mr. Jenkin's data base?
(b) Which columns in the database contain quantitative data?
(c) Which columns in the database contain qualitative data?
(d) Write down what Mr. Jenkin would put in his database if you joined his class.

2. Which of the following would give:

(a) qualitative data
(b) discrete quantitative data
(c) continuous quantitative data?

(i) Mass (ii) Number of cars
(iii) Favourite football team (iv) Colour of car
(v) Price of chocolate bars (vi) Amount of pocket money
(vii) Distance from home to school (viii) Number of pets
(ix) Number of sweets in a jar (x) Mass of chips in a packet
3. A traffic survey records information about cars passing a check point. Some data is given in the table below.

<table>
<thead>
<tr>
<th>Age of Car (years)</th>
<th>Colour</th>
<th>Speed</th>
<th>Number of Passengers</th>
<th>Trailer / Caravan</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Red</td>
<td>26 mph</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Blue</td>
<td>47 mph</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>White</td>
<td>36 mph</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>Red</td>
<td>31 mph</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Silver</td>
<td>33 mph</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>Green</td>
<td>29 mph</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>White</td>
<td>30 mph</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>1</td>
<td>Red</td>
<td>31 mph</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Blue</td>
<td>42 mph</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>Grey</td>
<td>28 mph</td>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

(a) Explain why the Number of Passengers is discrete data.
(b) Explain why Speed is continuous data.
(c) Which columns contain qualitative data?
(d) How fast was the silver car travelling?
(e) How many cars were towing a trailer or caravan?
(f) What colour was the slowest car?
(g) How fast was the car with the most passengers travelling?
(h) What was the age of the car with the highest speed?

4. The table below shows a database that has no entries.

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Favourite food</th>
<th>Favourite T.V show</th>
<th>Favourite pop group</th>
<th>Time spent watching T.V yesterday</th>
</tr>
</thead>
</table>

(a) You can add headings to the last two columns.
(b) Collect data from 10 people to complete the database.
(c) State whether each column contains
   (i) qualitative data;
   (ii) continuous quantitative data;
   or (iii) discrete quantitative data.

(d) Answer the following questions
   (i) What is the most popular T.V show?
   (ii) Who is the oldest?
   (iii) What is the favourite pop group for the youngest person?

(e) Write 3 more questions you could answer using your database and write the answers to them.

1.2 Questionnaire Design

The objectives of this section are to
   • understand how to design effective questionnaires
   • understand the difference between biased and unbiased questions.

Questionnaires must be designed carefully so that the answers given produce the required information: they should do so without influencing the people completing them. The following list contains points to consider when designing a questionnaire:

- The questions should be worded to provide the information needed by the researcher.
- Care must be taken not to invade people's privacy, so questions that do not relate to the purpose of the questionnaire must be excluded.
- The questions should be capable of being answered reasonably quickly.
- The questions should be easy to understand and should not be ambiguous.
- The questionnaire should not contain biased or leading questions.
- Questions may have possible responses presented in a multiple choice, or YES/NO format.
Where responses are provided, they should cover every possible answer.

The responses provided should not overlap.

The responses provided should not force people to answer in a way that they do not wish to answer.

The questionnaire should be designed so that the results are easy to analyse.

You should try out a questionnaire in a pilot study of a few people before using it with a large group of people. This will allow you the opportunity to alter questions that do not work well; for example, where they are misinterpreted.

Example 1

Anna wants to identify the favourite colours of children of different ages. Comment on the following questions that she has decided to ask:

1. **Which age range are you in?**
   
   0 - 5
   6 - 12
   12 +

2. **Please tick the colour that you like most from this list:**
   
   Blue
   Green
   Yellow
   Orange
   Black

Solution

- The questionnaire is easy to fill in and the data will be easy to collect and analyse.

- An adult could answer question 1 and you would not be aware of this, as the 12 + category would include children and adults.
• The age categories overlap. A 12-year old would not know whether to tick the second or third box.

• The survey only asks for the preferred colour from a limited choice. If you want to find the favourite colour, you will have to give many more choices (red and purple, for example, are colours which people might want to choose as their favourites).

• An alternative way of improving the second question would be to have an extra category labelled

   'Other colour (please specify)  ☐  . . . . . . . '

Example 2
Comment on the questions below:

1. Do cars cause pollution in the town centre? YES ☐ NO ☐
2. Do cars cause traffic hold-ups in the town centre? YES ☐ NO ☐
3. Are some car drivers a danger to pedestrians in the town centre? YES ☐ NO ☐
4. Do you think that cars should be banned from the town centre? YES ☐ NO ☐

Solution
The questions are biased. The first three are designed to focus on the disadvantages and dangers of cars, so that people are more likely to say 'yes' when they answer question 4.

Exercises
1. Design a questionnaire to find out whether people would be in favour of banning cars from your nearest town centre.
2. Design a questionnaire that could be used to investigate students' opinions of the method of transport that they use to travel to school.
3. Design a questionnaire to investigate how students rate the quality of the meals served in their school.

4. Design a questionnaire that can be used to determine whether the general public, in your area, would be in favour of a new youth club being built.

5. Rewrite the following questions so that they are not biased in any way:
   (a) Do you agree that maths is boring?
   (b) Are you in favour of town centre car park charges being increased in order to discourage car drivers from using their cars?
   (c) The price of a school meals has not increased for 2 years. Do you think that school meals are good value for money?

6. Comment critically on the following questions. In each case, rewrite the question to show the improvements you have made.
   (a) Are you young, middle-aged or old?
   (b) Please select your favourite breakfast cereal from this list:
       - Cornflakes
       - Porridge
       - Corn Pops
       - Froot Ooo’s
   (c) How old are you?
       - 0 → 5
       - 7 → 10
       - 12 +
   (d) Do you have any brothers?
       - Do you have one brother?
       - Do you have more than one brother?
       - Do you have at least two brothers?

7. Design biased questionnaires that would encourage people to reach the conclusion that the government:
   (a) dislikes car drivers,  (b) encourage car drivers.

8. The local council interviews people who are arriving in the town centre on a warm, sunny day. People are asked which of the following methods of transport they have used:
   - Walking
   - Cycling
   - Bus
   - Car
   (a) Explain why the results may not be reliable for deciding transport policies.
   (b) Suggest how the council should collect more data.
9. As part of a survey, school students were asked on which days they watched television during the previous week.
   (a) Describe what problems the researchers may have had in reporting on how much television these students watched.
   (b) Design a better question or questions to gather data for a report on how much television school students watch.

10. Design a questionnaire that could be used to gather data on how school students spend their summer holidays.

1.3 Data Collection

The objective of this section is to

- understand how to summarise data and use vertical line diagrams to display it.

Example 1

The marks below were scored by the students in a class on their maths test. The marks are all out of a possible total of 10 marks.

8  6  8  7  7
7 10 9  6  8
8  4  3  2  5
8  8  6  5  6
4  9  8  4  7
7  5  3  7  6

Draw a vertical line diagram to illustrate these data.

Use your diagram to answer the following questions:

(a) What is the most common mark?
(b) What is the highest mark?
(c) What is the lowest mark?
(d) What is the difference between the highest and lowest marks?

Solution

The first step is to organise the data using a tally chart, as shown below. This gives the frequency of each mark.

When drawing a tally chart, a short line is drawn for each of the first four times a particular answer is given, with a line drawn across the four lines for the fifth answer. This makes it easy for you to count up (or tally) the frequency of each answer.
The diagram can then be drawn as shown below. The height of each line is the same as the frequency; that is, the number of times it occurs in the data list.

(a) The most common mark is 8, which occurred 7 times.
(b) The highest mark is 10.
(c) The lowest mark is 2.
(d) The difference between the highest and lowest marks is $10 - 2 = 8$.

Note: a vertical line diagram is an appropriate way to represent information that consists of distinct, single values, each with its own frequency. A bar graph is more suitable for grouped numerical data.
Exercises

1. A teacher gives the children in her class a test, and lists their scores in this table:

   (a) Draw a vertical line diagram to illustrate these results.
   (b) What is the most common mark?
   (c) How many children are there in the class?

<table>
<thead>
<tr>
<th>Mark</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

2. The vertical line diagram below is based on data collected by a class about the number of children in their families:

   (a) What is the most common number of children per family?
   (b) How many children are there in the class?
3. The staff in a shoe shop keep a record of the sizes of all the shoes they sell in one day. These are listed below:

| 8 7 6 6 8 7 5 4 3 1 |
| 11 7 8 9 5 6 6 5 6 4 |
| 3 10 8 9 7 6 6 5 4 2 |
| 6 9 11 3 5 6 7 8 8 3 |
| 4 6 7 8 9 8 8 7 6 4 |

(a) Complete a tally chart for these data.
(b) Draw a vertical line diagram for these data.
(c) What advice could you give the shop staff about which size shoes they should keep in stock?

4. The figures below show the number of people entering a Post Office every minute during a period of half-an-hour.

| 4, 0, 1, 1, 3, 6, 2, 3, 3, 4, |
| 5, 0, 2, 3, 2, 4, 5, 2, 1, 1, |
| 3, 2, 2, 4, 6, 3, 3, 1, 3, 1 |

Complete a frequency table for this data.

5. Mr Graddon says that his class is better at tables than Mr Hall's class. The two classes each take a tables test, and the results are given below. The scores are out of 10.

<table>
<thead>
<tr>
<th>Mr Graddon's Class</th>
<th>Mr Hall's Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 6 7 8 9 10</td>
<td>4 7 8 3 5 6</td>
</tr>
<tr>
<td>0 1 3 6 9 2</td>
<td>7 4 5 6 6 5</td>
</tr>
<tr>
<td>5 1 2 2 0 1</td>
<td>5 5 6 7 4 3</td>
</tr>
<tr>
<td>6 4 0 1 10 9</td>
<td>4 5 6 6 7 8</td>
</tr>
<tr>
<td>1 2 3 5 10 9</td>
<td>6 7 5 6 4 5</td>
</tr>
</tbody>
</table>

(a) Use tally charts to draw vertical line diagrams for each class.
(b) Which features of the two diagrams would Mr Graddon use to support his claim that his class is better at tables?
(c) How would Mr Hall use the diagrams to argue the other way?
(d) Which class do you think is better at tables?
6. A gardener keeps a record of the number of Scotch bonnet peppers he picks from his plants during June. The number of peppers picked each day is listed below:

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>14</td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

(a) Construct a tally chart and draw a vertical line diagram for these data.
(b) What is the *largest* number of peppers picked on one day?
(c) What is the *smallest* number of peppers picked on one day?
(d) What is the number of peppers that was picked *most often*?

7. Mr Bryant keeps a record of the number of Grade 7 pupils absent each day over a 6-week period; his records are shown below:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

(a) Draw and complete a tally chart for these data.
(b) Draw a vertical line diagram for these data.
(c) What is the most common number of absent pupils?

8. A taxi driver writes down the number of times he has to stop at red traffic lights on all the journeys he makes one day; these are listed below:

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

(a) Draw and complete a tally chart for these data.
(b) Draw a vertical line diagram for these data.
(c) How many journeys did the taxi driver make that day?
(d) What was the most common number of red lights on these journeys?
(e) How many times in total did he stop at red traffic lights?