CHAPTER 7

DATA ANALYSIS AND VISUALISATION

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- sources of authentic data in large repositories
- factors influencing the integrity of data, for example accuracy, timeliness, authenticity, relevance
- characteristics of data types and data structures relevant to selected software tools

Approaches to problem solving

- types and purposes of data visualisations
- problem-solving activities related to analysing needs: functional and non-functional requirements and constraints
- characteristics of file formats and their suitability to be converted to other formats
- design tools for representing data visualisations
- formats and conventions applied to visualisations to improve their effectiveness for intended users
- functions of appropriate software tools to extract targeted data and to manipulate data when developing visualisations
- criteria and techniques for evaluating visualisations.

For the student

If you can imagine that the amount of data that is generated every day just by using social-networking tools, you might also be able to imagine that there is someone, somewhere, who is looking through a mountain of data looking for meaning.

Data visualisation is the process by which we take large amounts of data and process it into effective graphical representations that will meet the needs of users or clients.

These representations can take the form of charts, graphs, spatial relationships and network diagrams. In some cases, the data visualisation might involve interactivity and the inclusion of dynamic data that allows the user to deduce further meaning from the visualisation.

For the teacher

This chapter introduces students to the knowledge and skills needed to use software tools to access authentic data from repositories and present the information in a visual form.

The key knowledge and skills are based on Unit 2, Area of Study 2. If a data visualisation is effective, it reduces the effort needed by readers to interpret information. This chapter takes students through the different types of visualisations, and then uses a case study to explore some of the tools available to process data from repositories, such as the Bureau of Meteorology and the Australian Bureau of Statistics.

Information needs and data visualisations

When clients or users require particular information, and no system currently exists that provides the information, then an information need has been identified.

This could be due to an existing **information problem** (an organisation is worried about declining sales), an **identified need** (park rangers need a method to communicate weather conditions on total fire ban days), or an **opportunity** (currently no list of driving instructors in Victoria exists).

When an information need has been identified, one process used to help to create a solution that will meet the needs of the clients or users is the **problem-solving methodology (PSM)**.

The problem-solving methodology is discussed in Chapter 2.

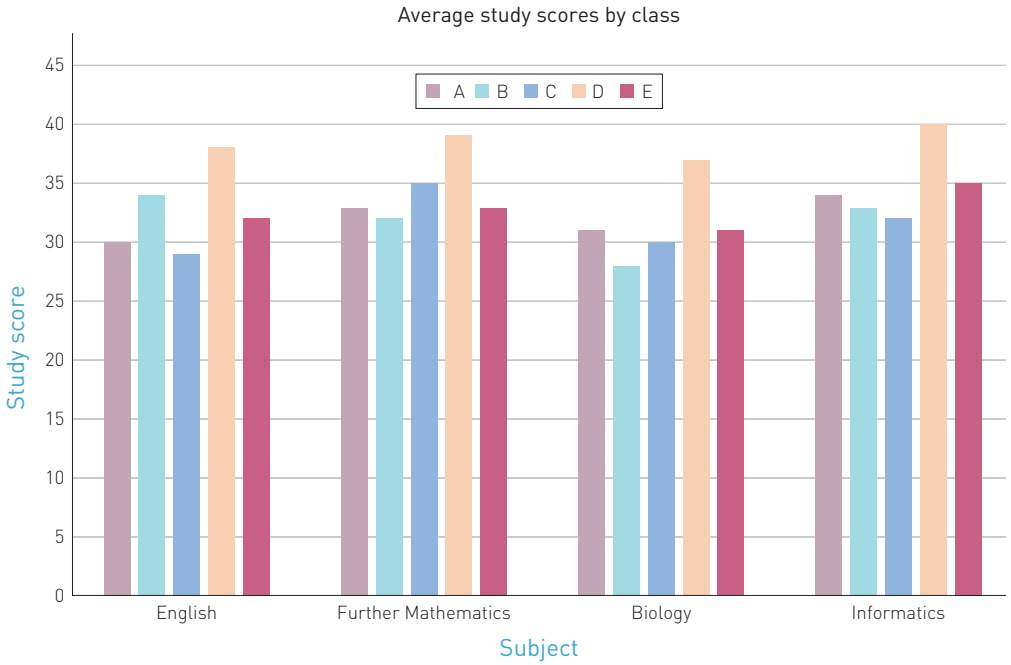


FIGURE 7.1 Visual representation of average VCE subject study scores by class

Sometimes the information need might revolve around finding a method to represent a data set. Looking at data in a table does not provide much meaning for the user. Table 7.1 displays the subject results for students in different classes.

TABLE 7.1 Student test results

Class	English	Further Mathematics	Biology	Informatics
A	30	33	31	34
B	34	32	28	33
C	29	35	30	32
D	38	39	37	40
E	32	33	31	35

At first glance, the data does not really have much meaning, nor are any relationships or patterns in the data identified. By following the four stages of the problem-solving methodology, the requirements needed from the solution will be identified, possible solutions designed, then

developed, and after the solution is created, the user can evaluate whether the solution is meeting their needs (the requirements identified in the analysis stage).

Figure 7.1 on the previous page is a solution that visually represents the data shown in Table 7.1.

The solution presented in Figure 7.1 used a column chart to represent the data. A column chart is useful for comparing different data items. In this situation the average study score of each class is compared with the other classes. When looking at the data in the table it is difficult to identify any patterns or relationships. Once the data is converted into a visual format it quickly becomes apparent that Class D performed the best in each of the subjects.

With the use of data visualisations, further investigation can be undertaken to understand why Class D performed better in the VCE than the other classes. Did Class D consist of all the high achievers at the school? Did they have the best teachers? Did the students in Class D complete more homework than the other classes, or was it just a coincidence?

THINK ABOUT COMPUTING 7.1

TV weather reports use data visualisations to present weather conditions. Think of other situations from day to day, where data is presented in a visual format that is more meaningful and is quick and easy to understand.

Sources of authentic data

As discussed in earlier chapters, there are two sources of data: primary and secondary.

Primary sources

Primary data sources involve collecting the data directly or firsthand from stakeholders. Methods used to collect primary data include **interviews**, **observations** and **surveys**. By its nature, data collected from primary sources should be considered authentic. Measures, though, need to be put in place when collecting from primary data sources to ensure the data is accurate, free from bias, reliable and relevant.

Secondary sources

Collecting data from **secondary data sources** involves using data that other users have collected firsthand. The range of sources holding secondary data include newspapers, books, magazines and reports, to name a few, with a wide range of organisations collecting data for their own particular information needs.

One important question to ask when using data from a secondary source concerns how authentic it is. Finding similar data from two or more sources helps to authenticate the data. Also, collecting data from organisations that are considered reliable and trustworthy helps to increase the chance that the data is authentic. **Government departments** collect and store vast amounts of data and take many precautions to ensure that the data they collect is authentic. They are increasingly making these data sets available for public use. As a result, government departments are considered a good source of authentic data. Some of these sources are discussed on pages 265–67.

Bureau of Meteorology

The Bureau of Meteorology (BOM) is Australia's national weather, climate and water agency. They collect a wide range of climatic data to be able to make regular forecasts, issue warnings and offer advice. Much of the data that they collect is available for public use.

Bureau of Meteorology

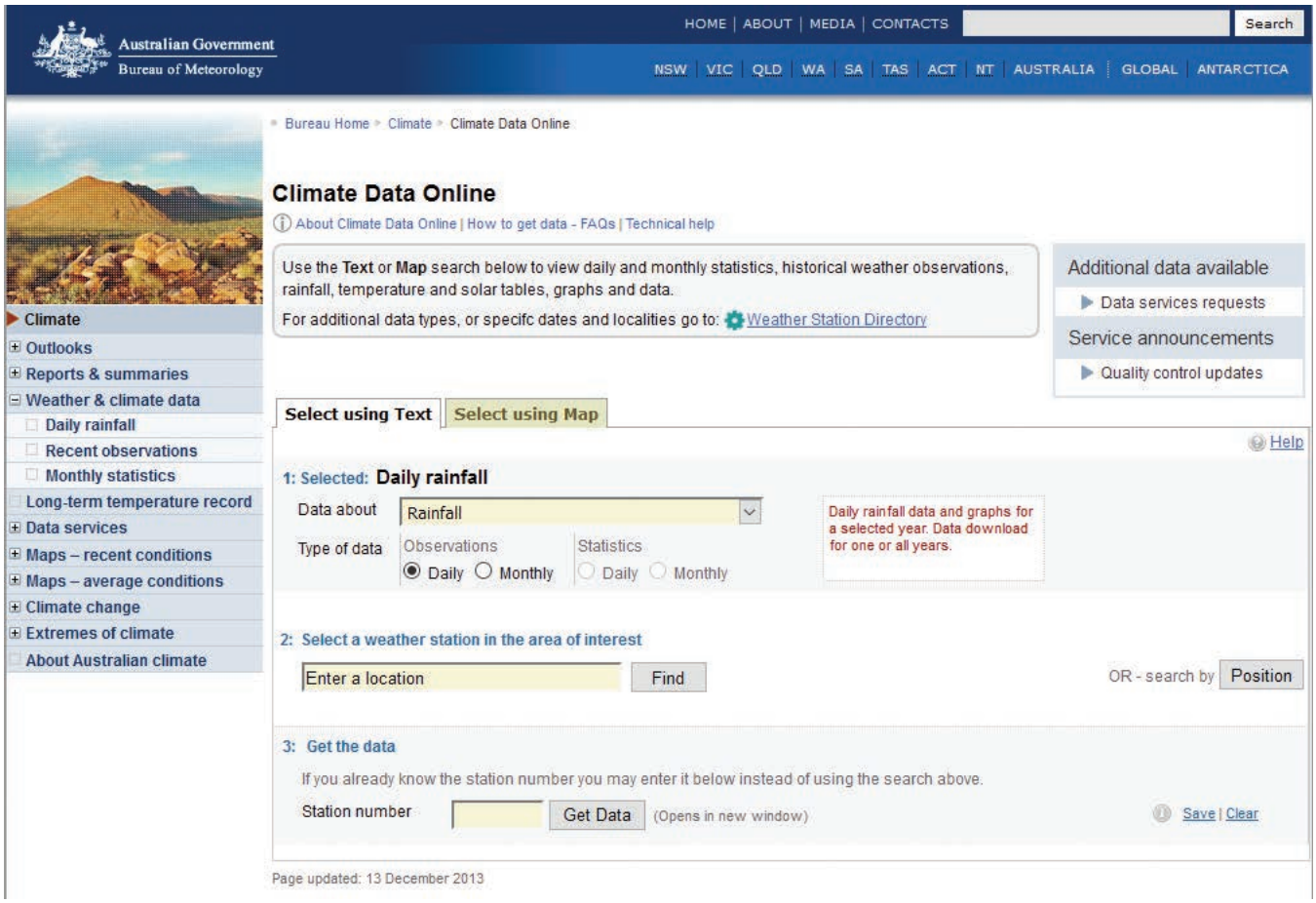


FIGURE 7.2 A screenshot from the BOM's website

Victorian government data directory

The Victorian government data directory contains a range of data sets from Victorian government departments. There are data sets available on a diverse range of subject areas including education, finance, health and town planning. The data sets offered are in a range of file formats, including spreadsheets, audio and video, comma separated values, geospatial and API tools.

Victorian government data directory

https://www.data.vic.gov.au

FIGURE 7.3 A screenshot from the Victorian government data directory website

Australian Bureau of Statistics

The **Australian Bureau of Statistics (ABS)** is the statistical agency of the federal government. The ABS provides statistics on a wide range of economic, environmental and social issues, for use by governments and the community. Data sets are available on a diverse range of topics from foreign trade, agriculture, sporting facilities to crime statistics. In fact, the ABS website provides an endless supply of open data.

ABS

FIGURE 7.4 A screenshot from the ABS website

Data.gov.au

Data.gov.au is similar to the Victorian government data directory in that it provides publically available data sets that have been collect by federal government departments.

Many other organisations offer data sets to the public.

Data.gov.au

FIGURE 7.5 A screenshot from Data.gov.au

Data integrity

Data integrity refers to the quality of the data. An information system's purpose is to transform data into useful information. It therefore is important that the data inputted has integrity, because the information produced is based on the data itself. For example, if a customer database contains the wrong postal details for a customer this is incorrect data. When a data set contains these types of errors, it loses its integrity. The more errors the data set contains, the lower its integrity.

There are many ways the quality or integrity of the data can be measured. When using data from either primary or secondary sources, factors to consider include accuracy, timeliness, authenticity and relevance.

THINK ABOUT COMPUTING 7.2

Create a list of four other organisations in Australia who provide data sets for public use. Each organisation can be a government department or private organisation.

Accuracy

Accuracy involves ensuring that the data collected is correct and does not contain any errors.

When using primary methods to collect the data, validation may be able to be used to reduce the chances of incorrect data being entered. Data validation often involves restricting the data that can be entered into a particular field and by restricting what can be entered. The chances of incorrect data being input are therefore reduced.

Many online forms contain a number of validation methods that help reduce the chances of errors being input. Validation techniques include dropdown lists, radio buttons, predictive text, check boxes and required fields.

THINK ABOUT COMPUTING 7.3

How many validation techniques are used in the online checkout form from cafe press (Figure 7.6)?

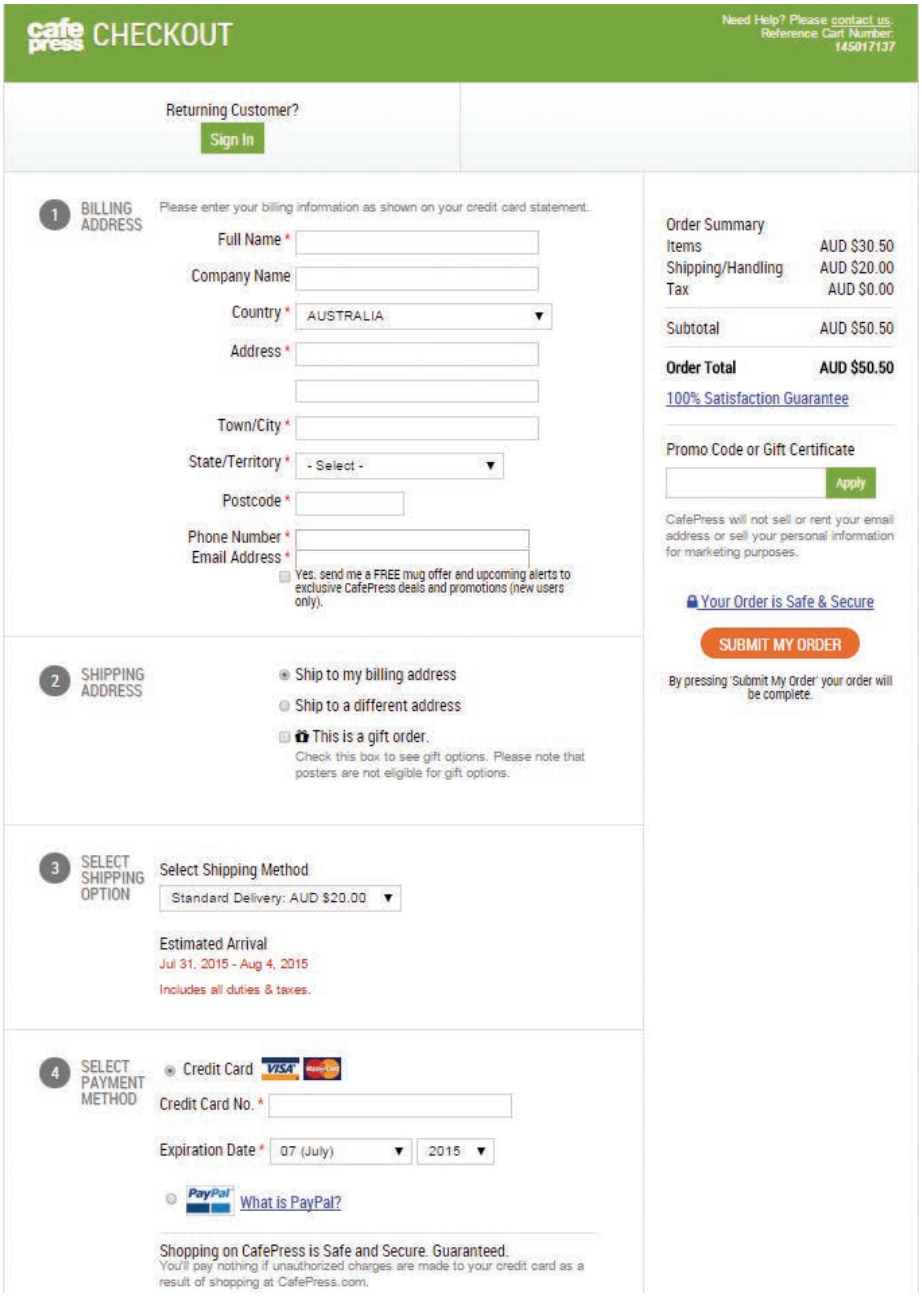


FIGURE 7.6 Online form containing validation techniques

Timeliness

Timeliness relates to the age of the data – how old is it? The data used should be relevant for the time period. For example, using Melbourne's population data from the 1990s to help plan the location for new primary schools would result in a plan that does not match the city's current needs. It is important that the data input into the information system is timely to match what is needed, and is not collected too early (or too late).

Timeliness can also relate to the information provided by the information system. If the information produced is not provided in a suitable timeframe it would be useless. Imagine a school produces a daily bulletin outlining all the events that occur each day, but the bulletin is always published at the end of the school day, and not the start. The information is not being received in a timely fashion and, therefore, is useless to the organisation.

Authenticity

Authenticity relates to how genuine the data set is. Is the data genuine, original, accurate, reliable and trustworthy? The authenticity of primary data is easier to confirm as it has been collected firsthand, although precautions should be put into place to ensure it is accurate and reliable. Also, when collecting primary data it is important that the sample size is not too small, as this may lead to inaccurate results.

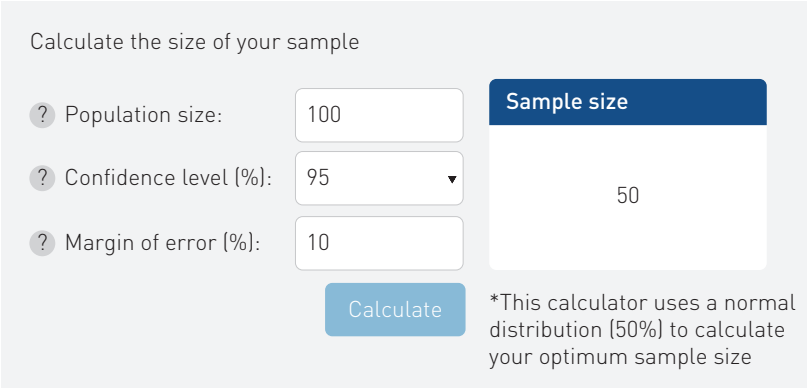


FIGURE 7.7 Online calculator to calculate optimum sample size

Relevance

Relevance is related to whether the data helps to produce the information required to meet the information need of the user. There is no point using the data that does not contribute to producing the information required. Collecting data about the weather is useless if the purpose of the system is to identify where new hospitals should be located in Australia. There are many attributes of relevant data but ultimately the data should help achieve the needs of the user.

Data types and data structures

Data types are particular forms that an item of data can take, including numeric, character and Boolean. The number of data types can vary between software applications. Each piece of data can be classified into one of the data-type groups. Common data types include the following.

Character

A **character** data type holds a single letter, number or symbol and is usually stored in 2 bytes of memory. Numbers stored as a character data type cannot be used in mathematical calculations. Examples of characters include `a`, `&`, and `5`.

Text (string)

A **text** data type consists of a 'string of characters'. The terms 'text' and 'string' are sometimes used interchangeably to refer to the same data type. Within software applications, the term 'text' is often used. If the data consists of 2 or more characters then a string data type is used. The amount of memory used to store the string will depend on the length (2 bytes × length of string). Again, numbers stored as a string cannot be used in mathematical calculations. Often a phone number is saved as a string. Examples of strings are `Hello`, `0406000000` and `ABC123`.

Integer

An **integer** is a number without a fractional or decimal component, e.g. a whole number. An integer is usually stored in 4 bytes of memory. Numbers stored as integers can be used in mathematical calculations. Examples of an integer include `-345`, `0`, and `1234567890`.

Floating point (or decimal)

A **floating point** number is a number with a fractional or decimal component, such as a decimal number. A floating point number is split into two parts: The mantissa and the exponent. The mantissa holds all of the digits with a decimal point after the first digit. The exponent holds the power of 10 by which the mantissa must be multiplied to regain the original number. A floating point can take up to 12 bytes of memory. Examples of floating point data types include `3.17`, `-6.890` and `0.001`.

Boolean

A **Boolean data type** can only hold two possible values, usually true or false. A Boolean data type is known as a logical data type. It is often used in conditional statements to test if a condition or value is true or not. Boolean data types are efficient in terms of storage space as they only require 1 byte of memory for storage.

THINK ABOUT COMPUTING 7.4

In theory, a Boolean data type only requires 1 *bit* of memory space. Find out why it needs to take up 1 byte of memory instead.

Data structure

A data structure is a particular way of organising a collection of data items. Using data structures helps to improve the efficiency of the data handling process. Three common data structures are arrays, records and files.

Array

An **array** is a collection of data items generally of the same data type. Each item in the array is allocated an address.

Table 7.2 illustrates an array data structure. The structure is named 'Number' and holds six data items – all integers. The value of Number (0) is 40 and Number (5) is 67.

TABLE 7.2 Array data structure

Number	
Index	**Value**
0	40
1	34
2	56
3	21
4	12
5	67

Record

A record is a collection of data items of different data types. Each element of a record is generally called a field. The record in Table 7.3 consists of six fields and involves data of different data types.

TABLE 7.3 Record data structure

Customer ID	Name	Phone	VIP	Quantity	Cost
01901	Karen	0406348790	Y	5	3.99

THINK ABOUT COMPUTING 7.5
Referring to Table 7.3, how many different data types can you identify?

File

A file data structure can hold numerous data items, arrays or records. A file is saved separately from the software program that utilises the file. Large data sets can be read from, or written to, a file for storage purposes. The file to the right shows a collection of records related to student results.

Figure 7.8 is an example of a comma-separated value (CSV) file. A range of file types will be discussed later in the chapter.

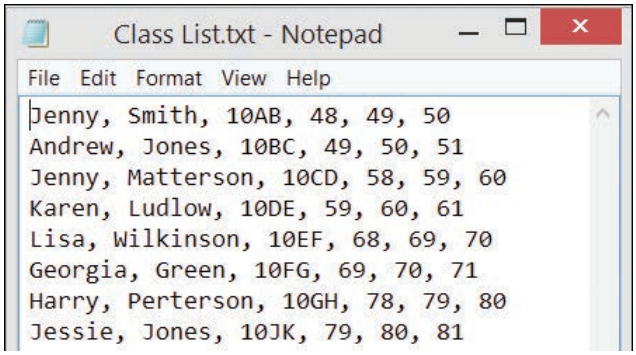


FIGURE 7.8 File data structure

Types and purposes of data visualisations

Data visualisation is a relatively new field of computing. As more and more data sets are being made available, the number of individuals and organisations performing analysis on this data is also increasing. The types of visualisations being used is endless. Data is being manipulated and processed into new forms of visualisations every day.

Figure 7.9 shows different types of charts used for different purposes, but these only scratch the surface in terms of types of data visualisations.

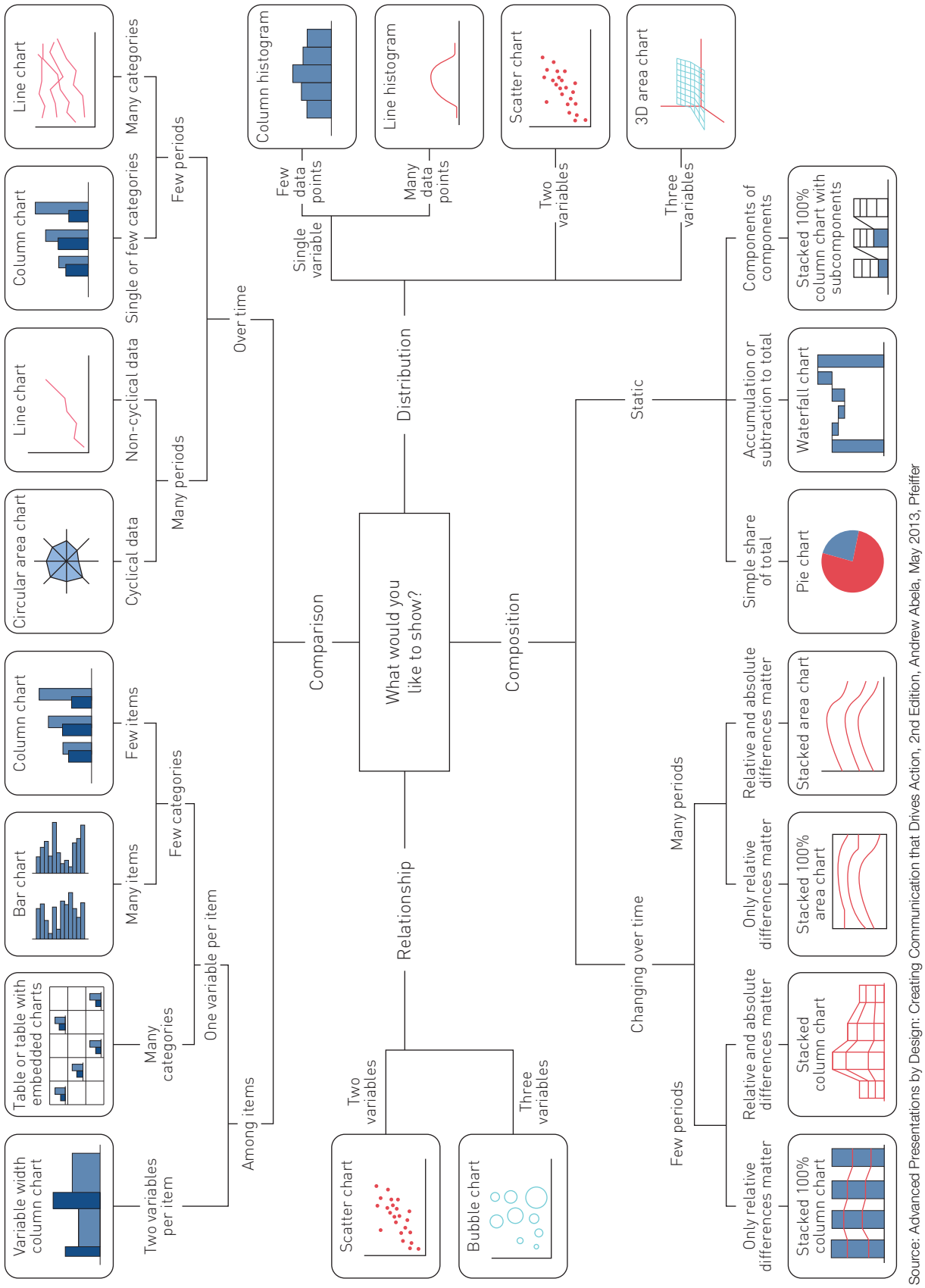


FIGURE 7.9 Some common types of data visualisations; it does not include network diagrams or maps

Charts

A **chart** (also known as a graph) is a method of displaying data visually, where the data set is represented as symbols in the chart. Charts are a popular way of presenting data visually and many spreadsheet applications have charting (or graphing) capabilities. Features of a chart can include a title, axis, scale or grid, data labels and a legend.

Charts are often used to visualise numerical data. There are a range of chart types, where each type is used for different purposes. A bar chart can be used to compare different items, a pie chart to show each data item as a proportion of the population, line charts are useful for showing the trend in a data item over time and histograms are useful for grouping data then showing the frequency of each group.

Shutterstock.com/vasabii

FIGURE 7.10 Types of charts

Maps

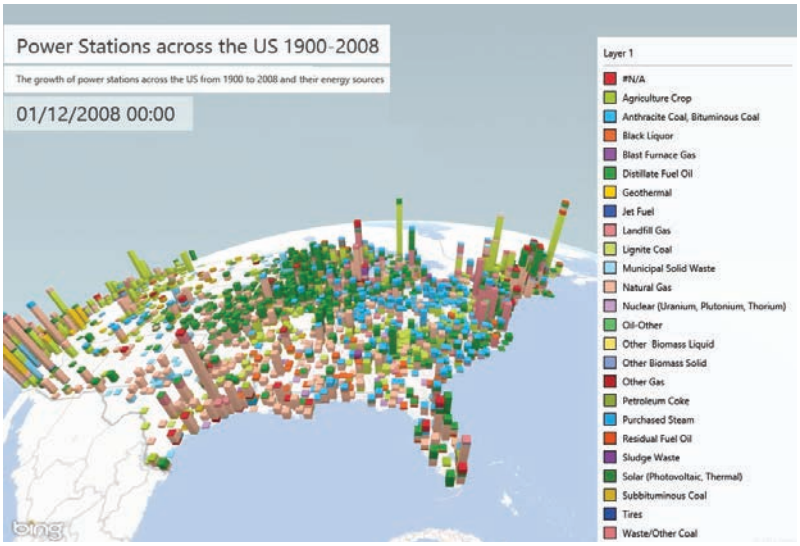


FIGURE 7.11 Geospatial visualisation

A popular method to display geographical data is by using **map-based visualisations**. These types of visualisation are often called **geospatial visualisations**. Geospatial data is data that is related to the geographical location covered. Data could be related to population, roads, rivers, climate, mobile phone towers or any other characteristic of the area. Many geospatial visualisations are dynamic and allow the user to zoom in or out or navigate over an area. Geospatial visualisations are becoming more popular because they are a very powerful tool that allows the data to be brought to life through visualisation. As a range of data can be overlayed with a geographical location the uses of these types of visualisations are enormous. Common uses have been for agricultural, environmental, mining and urban planning purposes, but the list is endless.

Networks

Network visualisations show relationships between different data items and relationships between different data sets. Finding relationships within and between data sets has been an increasing area of interest in recent years as more data has become publically available from both government and private organisations around the world.

A network visualisation might show the frequency with which individual players might pass the ball to each other in a football game or the number of people who travel on a public transport system each day. Network visualisations are also used to represent the layout of computer networks, or public transport systems. Figure 7.12 shows an example of a network visualisation.

NelsonNet additional resource: Figure 7.11 Geospatial visualisation, p. 274

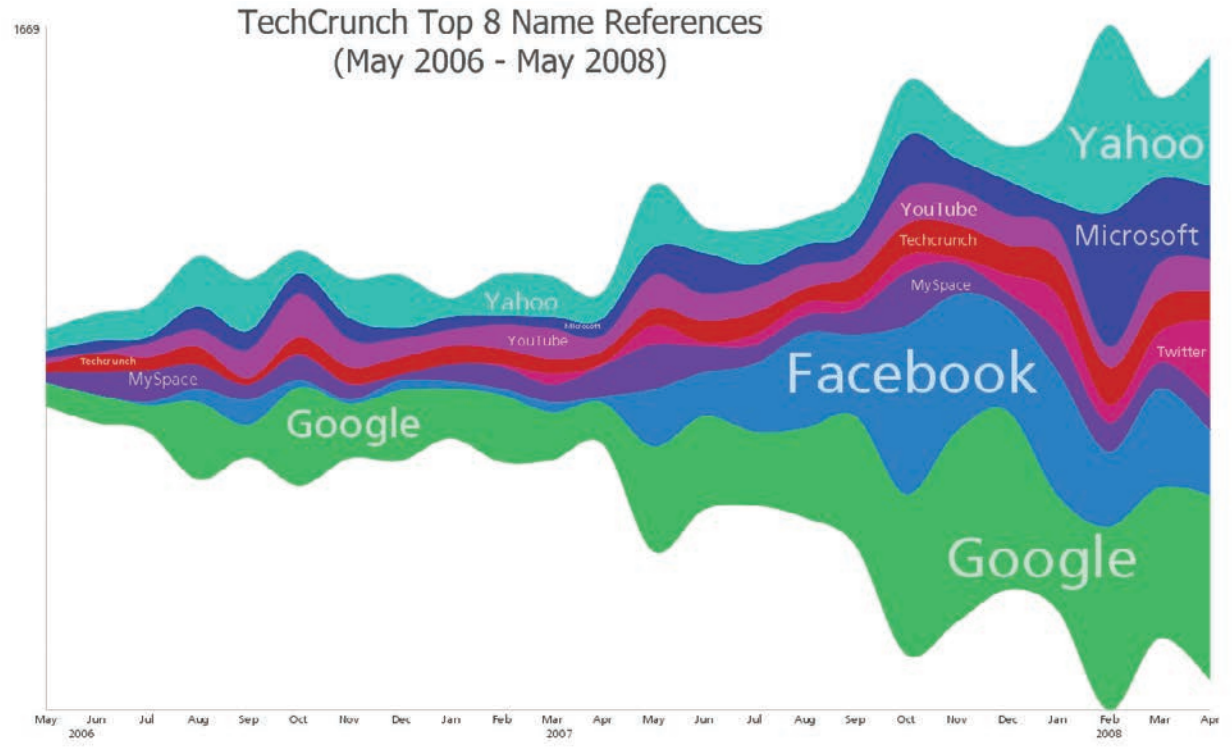
Alamy/Route55

FIGURE 7.12 Network visualisation

Time series

Time visualisation represents a data item or data set over a period of time. Some time-based visualisations will show historical data, while others capture live data to provide real-time information. It is also possible to display the dimension of time by adding motion or animation to create a dynamic data representation.

The data could also be related to timeline or time series. Timeline data may relate to individual items or events and show the order in which the items or events occurred over a time period, while time series data may relate to the same data item and show the variations or changes in the item over a time period. The visualisation shown in Figure 7.13 is an example of a time series representation, showing the top eight name references to social networking sites on the internet from 2006 to 2008.



Jeff Clark, Neoformix.com

FIGURE 7.13 Time visualisation

Hierarchy

Hierarchy visualisations show the relationships and structure between data items. In a hierarchy, data items are represented as being above, equal or below other items in the data set. Types of hierarchy visualisations can include organisational, structure and tree charts and mind maps.

They can be used to show the relationship between items in a data set or they can illustrate the breaking down of items into smaller components. Hierarchy visualisations are also useful for representing non-numerical data types. Figure 7.14 shows a word map representing Chapter 3 of this book. The subject networks have been broken down into sections and then the components of each section branch out from the centre.

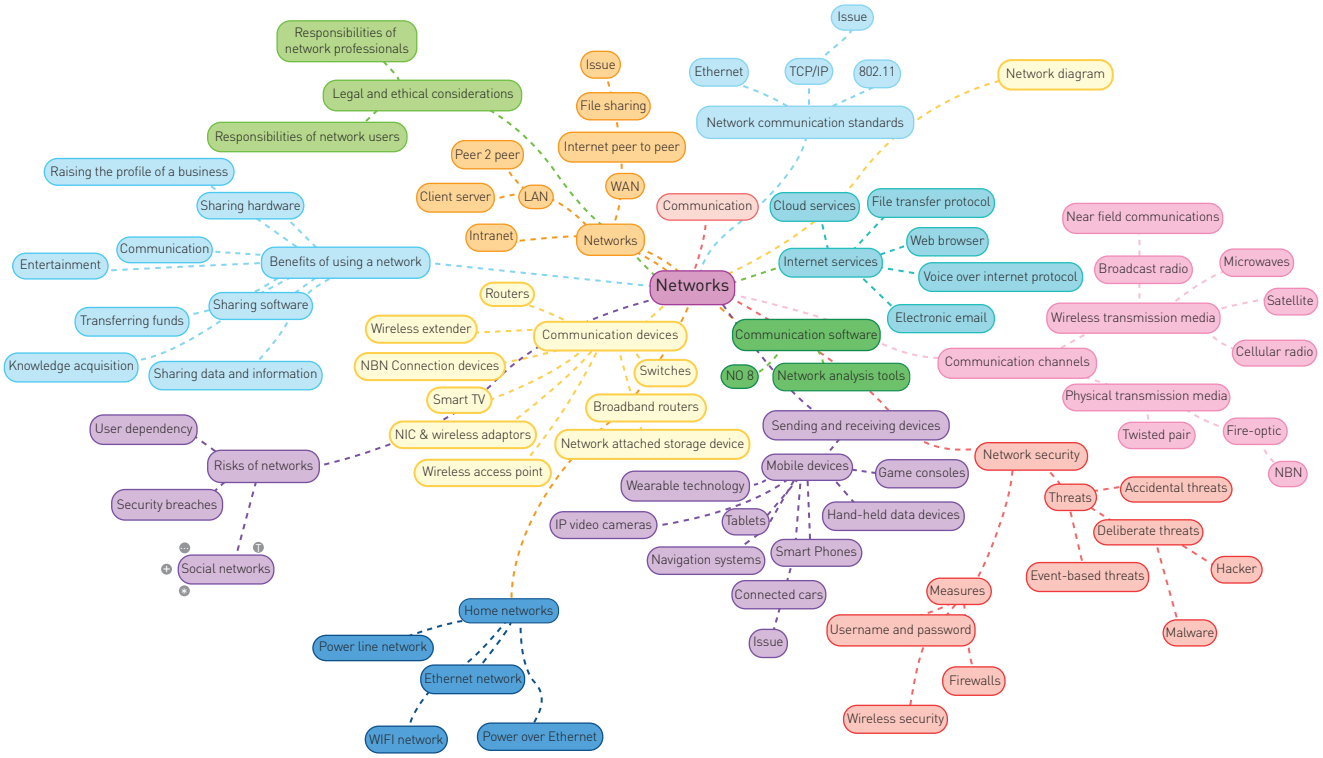
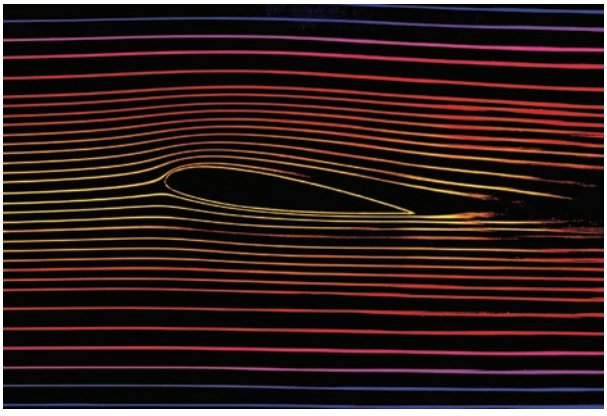


FIGURE 7.14 Hierarchy visualisation

Flow

Flow visualisations involve representing data that illustrates the flow pattern of a data item or items. This could be the pattern of customer movements through a supermarket or the series of pages a user would visit on a website to complete a transaction (user-flow diagram).

THINK ABOUT COMPUTING 7.6
User-flow diagrams are studied in Unit 3, Outcome 1 Informatics. Using a search engine, find three examples of user-flow diagrams.



Science Photo Library/Omikron

FIGURE 7.15 Flow visualisation

Flow visualisations are also used for scientific purposes to visualise the flow patterns of objects that are normally invisible, including air and water. Figure 7.15 represents the effect of an aircraft wing on the airflow passing the wing. The data for this visualisation was collected during testing using a wind tunnel, and the data converted to a visualisation.

Matrix

Matrix visualisations can be used to show the composition of individual items in the sample size. In this regard, they can be considered similar to pie charts. Matrix visualisations often divides the display area up into grids (similar to cells in a spreadsheet). Different sections of the display area are then used to represent the proportion of individual (or groups) of data items.

Matrix diagrams can be used to compare different data items, similar to a scatter chart. Again, the display area is divided in a grid or table format. Figure 7.16 shows a number of matrix visualisations with each visualisation representing a county in the United Kingdom. The colours represent the percentage of people who voted for a particular political party in an election.

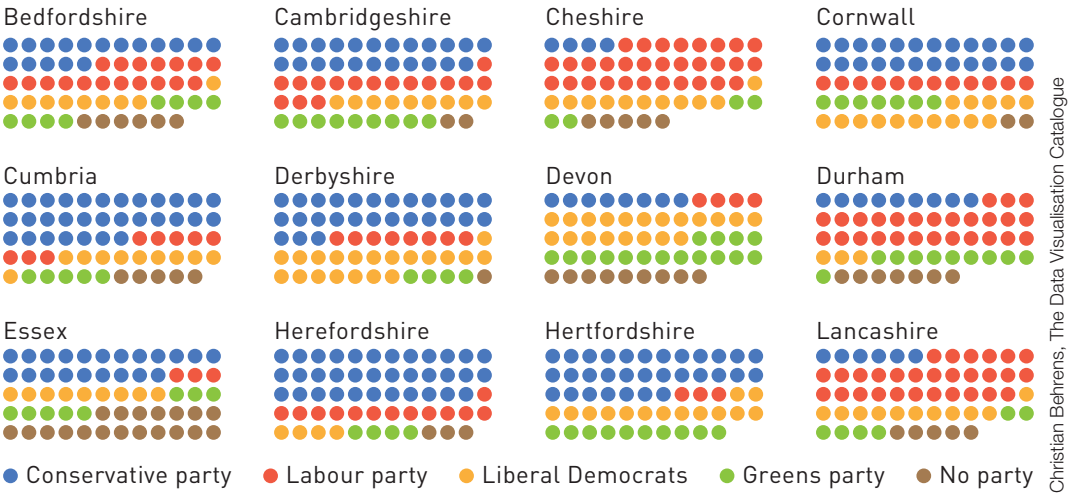


FIGURE 7.16 Matrix visualisation

PSM: Analysis

As mentioned earlier, when an information need is identified, one method to create a solution that will meet the need is to use the problem-solving methodology.

The first stage of the methodology, the analysis stage, is often considered the most crucial. Many organisations invest more time, effort and money for this stage, as they consider getting the analysis stage correct and having a clear picture of what is required will save time and money in the latter stages.

The analysis stage is typically about 'what' questions. What is the current information need? What is required to meet the current information need? What constraints may restrict the requirements? What is the scope of the solution?

The analysis stage consists of three activities: determining solution requirements, determining solution constraints and deciding upon the scope of the solution.

THINK ABOUT COMPUTING 7.7
Explain how completing a thorough analysis stage may save time and money in the latter stages of the methodology.

Solution requirements

Solution requirements are what the client needs from the solution. What features do they want in the solution? Solution requirements can be broken down into functional and non-functional requirements.

Functional requirements are directly related to what the solution will do. Some examples of functional requirements for data visualisations include:

- display population data of towns in Victoria in a visual format
- represent the distances between these towns
- allow the user to zoom in on a particular region or town
- allow the user to select an individual town to read more detailed information.

Non-functional requirements are other requirements that the user or client would like the solution to have but that do not affect what the solution does. Examples include:

- user friendly
- not display any personal details
- compatible with different web browsers and operating systems
- portable so it can operate on devices of different sizes.

When buying a car the amount of money available for the purchase may restrict the user's requirements. The user may not be able to purchase the car they really want. Instead they may need to re-evaluate their needs.

Solution constraints

Solution constraints are factors that may limit or restrict the solution requirements.

Like requirements, constraints can be broken down into groups: economic, technical, social, legal and useability. Chapter 6, Figure 6.7, page 229, shows an example of each type of constraint.

Economic

Economic constraints include time and cost.

The deadline by which the user or client needs to have the solution operational will define the time available to design and develop the solution. The longer the time available, the more time there is to complete an in-depth analysis, detailed designs and develop advanced features of the solution. The shorter the timeframe, the faster that each stage in the problem-solving methodology needs to be completed.

Meanwhile, the funds (money) available to complete the project may affect the hardware and software (digital systems) available for use, the number and range of staff who are available to work on the solution and even the data used as input, if the data sets required need to be purchased.

Both a lack of time or money may result in a re-evaluation of the user's requirements, or a re-evaluation of how the requirements can be achieved.

Technical

Technical constraints are constraints related to the hardware and software available for the project. Available hardware and software, memory and storage capacity, processing and transmission speeds, and security concerns are all examples of possible technical constraints.

THINK ABOUT COMPUTING 7.8
List three other technical constraints developers of smartphone apps need to consider when developing a product.

For example, developers need to keep in mind that smartphone users may not always have access to a high-speed network connection, so they need to ensure that any animated data visualisation solution does not require a large amount of bandwidth to download and view.

Social, legal and useability

Non-technical constraints relate to areas other than hardware and software. Useability and the user's level of expertise (social) are examples. If a solution is being developed for users with little digital systems expertise, this may restrict some of the requirements that would involve complex manoeuvres to complete. Creating a solution for a child audience may restrict the method used to input data into the solution.

Legal requirements are another type of non-technical constraint. Privacy laws may restrict features linked to displaying personal data in the solution, or to collecting data from the devices of someone using your solution. Copyright laws may restrict features that allow other users to upload content to the solution.

Scope of solution

The scope outlines the boundaries or parameters of the solution so all stakeholders are aware of exactly what the solution will contain. The scope of the solution consists of two elements. Scope is not required for your solution for this Area of Study.

Many house and land package contracts state exactly what is included with the package and what exactly is not included in the package for the price. For example, tiles on the floor of the kitchen are included, but the garden will not be landscaped.

What the solution will do

What the solution will do is a list of all the solution requirements (both functional and non-functional) that will be included in the solution.

What the solution will not do

This is a list of all the solution requirements that will not be included in the solution.

Usually these are solution requirements initially sought by the client, but because of constraints they have been left out of the solution project.

At the start of the project, it helps to outline what will and will not be included in the solution to prevent arguments later in the project between the client and the developer.

THINK ABOUT COMPUTING 7.9
What problems do you think a clear scope of solution can avoid later in the project?

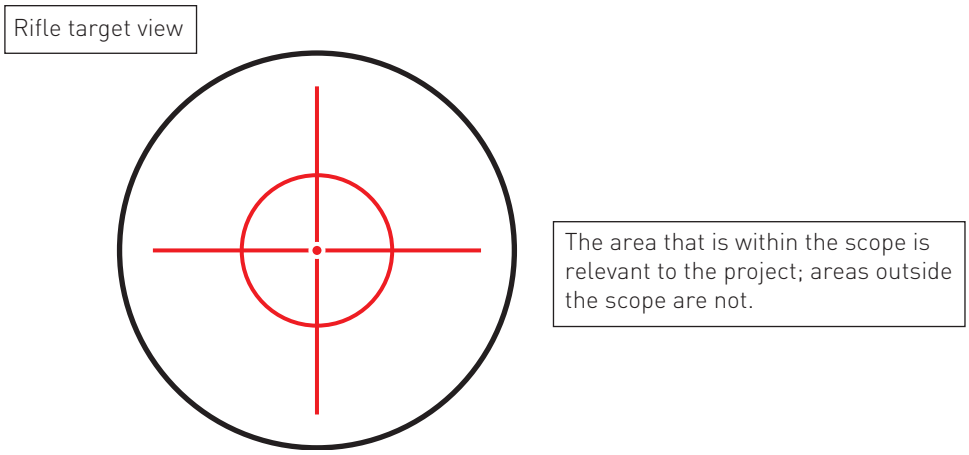


FIGURE 7.17 Scope of solution

An example of a scope of solution would be as follows: The solution will display population data of towns in Victoria in a visual format and graphically represent the distances between those towns. It will be created to be user friendly and for privacy reasons, it will not display any personal details.

The requirement that allows the user to zoom in on a particular region or town will not be included in the project because of economic factors, but may be added at a later stage.

Design tools

The second stage of the problem-solving methodology involves designing the solution. The analysis stage was all about 'what' the solution required. The design stage is all about 'how' the solution will be achieved.

The two activities involved in the design stage are the solution design and evaluation criteria. We will look at evaluation criteria in a later section of the chapter.

Solution design

The solution design involves planning how the solution will appear and function. To assist with planning, a range of design tools can be used to represent the solution's appearance and functionality.

Appearance design tool – layout diagram

A layout diagram can be used to plan how the visualisation will appear. The diagram can be hand drawn or computer generated, but if computer generated, should not use the same software as the solution. The purpose of a layout diagram is to illustrate how the final solution will appear. Once completed, the layout diagram can be inspected to see if the type of visualisation chosen is appropriate for the data set.

A number of alternative layout diagrams can be created to represent the data set in a variety of types of visualisations, to help decide which type to use in the final solution.

The layout diagram allows the developer and the client to give feedback on other areas of the solution, including formats and conventions used, colour schemes and other design principles. It is much easier to make changes to the solution in the design stage than once development begins.

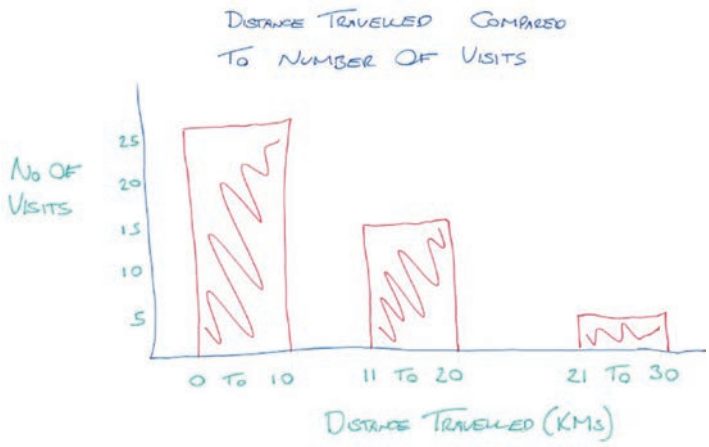


FIGURE 7.18 A hand-drawn layout diagram set out as a column graph

Nisha Samuel Design
Scroll to the section at the bottom of the page and click to view the 'Crowdfunding' storyboard.

Appearance design tool – storyboard

Storyboards can be used to show how the data-visualisation animation might work. For example, if you are using animation in the data visualisation, the storyboard might represent different stages of the solution. View Nisha Samuel's website for an example of a storyboard showing the sequence of scenes in a visualisation used to promote the understanding of crowdfunding communities. The website also shows other useful storyboards.

IPO charts are discussed in Chapters 2 and 3.

Functionality design tool – IPO chart

An input-process-output (IPO) chart is a design tool that helps to plan how a solution will function. The IPO assists in identifying the data required to produce the information (or output) required by the solution. Often by reverse engineering from the output, you can identify what processing steps are needed for the input data.

TABLE 7.4 Input process output design tool

Input	Process	Output
Distance travelled Number of visits	Select data Select insert tab Select column chart Add title and labels	Column chart

Functionality design tool – flowchart

Flowcharts can be used to show the procedure users follow to create a data visualisation.

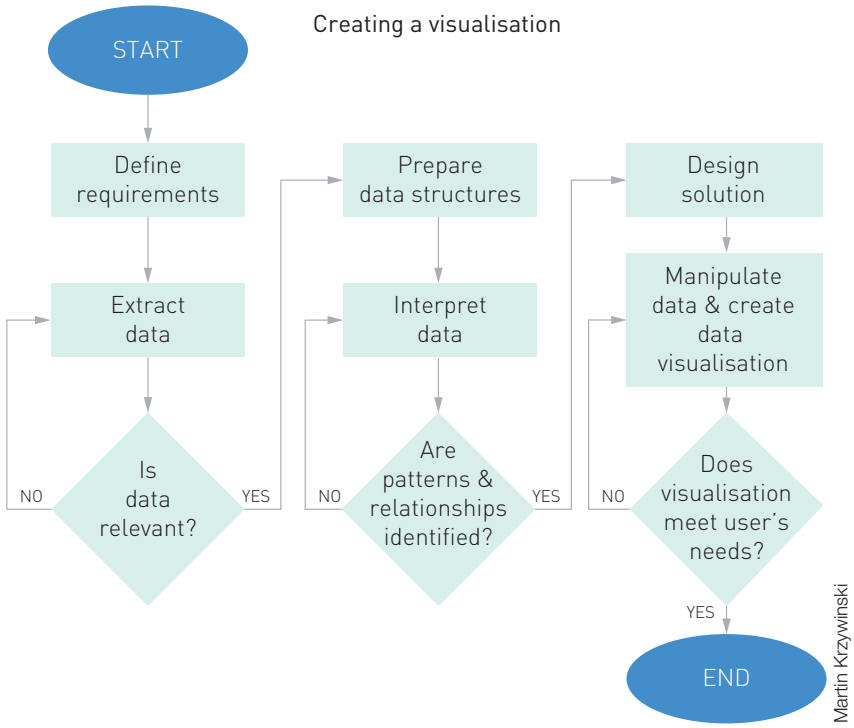


FIGURE 7.19 Flowchart on how to make a visualisation

Formats and conventions

Formats are related to the font type and size selected, background colour used and any other change of appearance. Conventions are general rules that are followed when using a particular format. When we present data in a table, often the column headings are in bold, the text left aligned and the numbers right aligned. When presenting data in a website, we use clearly named hyperlinks to make it easy for the user to find the required information. When addressing a letter, we often follow standard rules (or conventions). For example, the stamp goes in the upper right-hand corner, the return address goes in the upper left-hand corner while the 'to' address is centred on the envelope.

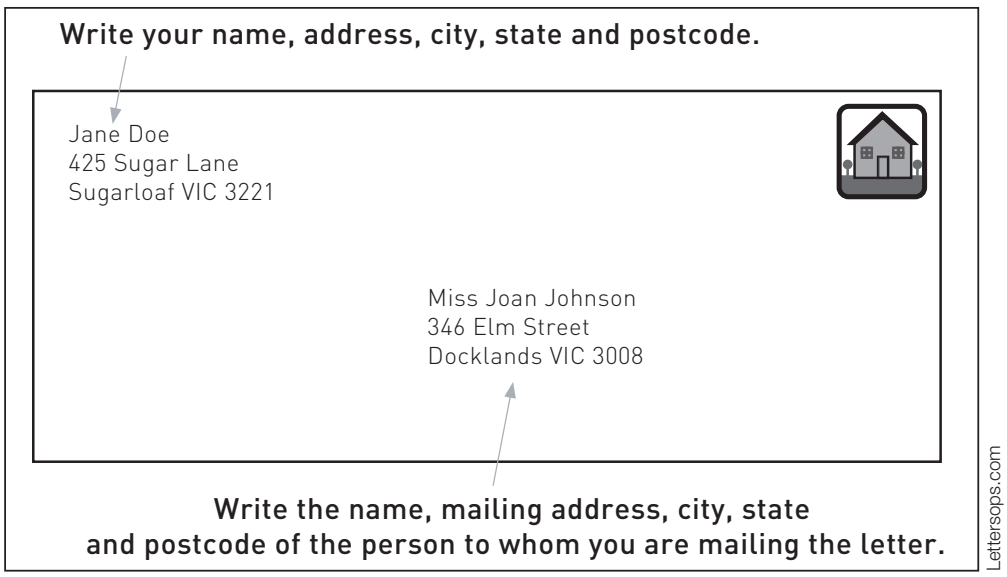


FIGURE 7.20 Conventions followed when addressing an envelope

With data visualisations appearing in so many different formats, there really is a range of conventions that would apply to each format.

Some common conventions that would improve the effectiveness of the visualisation would include:

- clearly title the visualisation and explain its purpose
- label axes if appropriate
- use a key or legend to identify different data items
- include the name of the author and the source of the data
- identify the units of measurement
- choose colours that match the information being discussed.

THINK ABOUT COMPUTING 7.10
Study a number of data visualisations you find effective. Try to identify the formats and conventions that have been used to make the solution stand out.

Software tools and functions

For this Outcome, you should follow two discrete steps to create your data visualisations. The first step involves extracting data from data sources. The second step involves presenting the data using a data visualisation.

Extracting data

Spreadsheet software

Many data sources provide data in a file format that is compatible with existing spreadsheet software applications.

The image below shows a data set available from the Victorian Government website Data.gov.vic.au, containing all of the details and locations of AFL teams and programs in Victoria. The data set is available in CSV file format that is compatible with spreadsheet software (Figure 7.21).

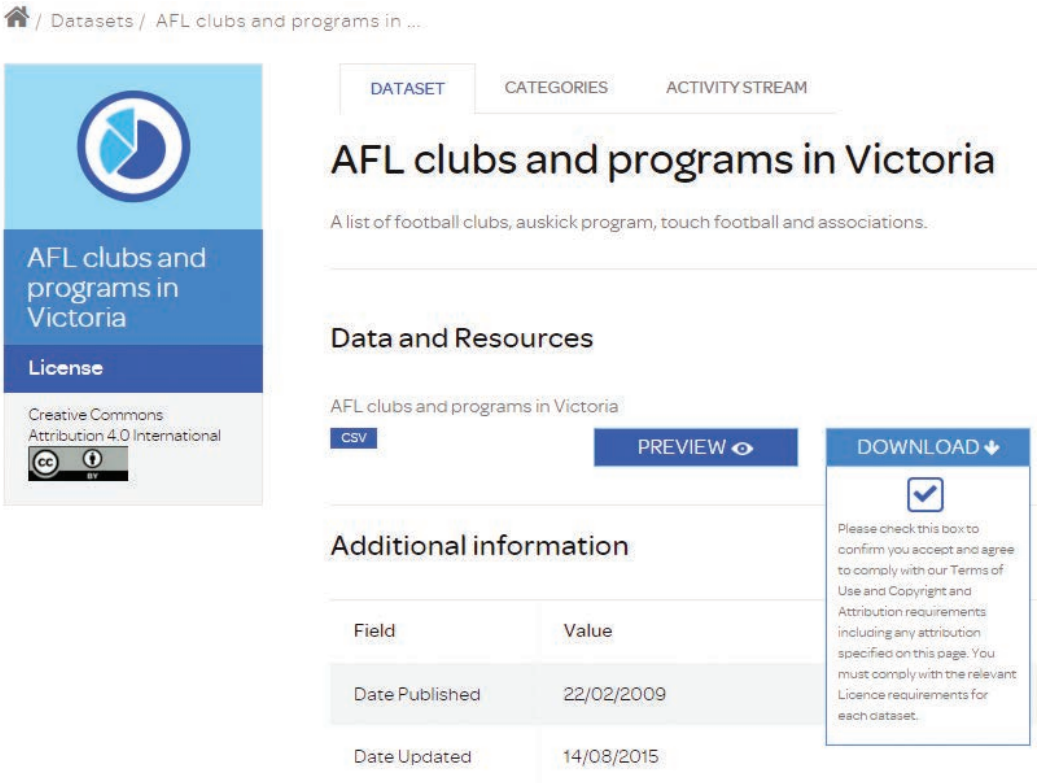


FIGURE 7.21 Data about AFL clubs and programs in Victoria

	A	B	C	D	E	F	G	H	I
1	Name	Address	Suburb	Postcode	State	Category	LGA	Region	
2	AFL Auskick - Aberfeldie	Aberfeldie Primar	Essendon	3040	VIC	AFL Auskick	Moonee Valley (	North and West Metropolitan Region	
3	AFL Auskick - Airport West	Hansen Reserve	Airport West	3042	VIC	AFL Auskick	Moonee Valley ;	North and West Metropolitan Region	
4	AFL Auskick - Ajax	Princes Park	Caulfield South	3162	VIC	AFL Auskick	Glen Eira	Southern Metro	
5	AFL Auskick - Alexandra	5665 Maroondah	Alexandra	3714	VIC	AFL Auskick	Murrindindi	Hume	
6	AFL Auskick - Alfredton	Alfredton Reserve	Alfredton	3350	VIC	AFL Auskick	Ballarat	Grampians	
7	AFL Auskick - Allansford	Allansford Recrea	Allansford	3277	VIC	AFL Auskick	Program		
8	AFL Auskick - Alphington/Fa	irfield	Northcote	3070	VIC	AFL Auskick	Yarra ; Darebin	North and West Metropolitan Region	
9	AFL Auskick - Altona	Altona Football C	Altona	3018	VIC	AFL Auskick	Hobsons Bay	North and West Metropolitan Region	
10	AFL Auskick - Altona Meac	Comden Reserve	Altona Meadows	3028	VIC	AFL Auskick	Wyndham ; Hob	North and West Metropolitan Region	
11	AFL Auskick - Amateurs	Queens Park Oval	Newtown	3220	VIC	AFL Auskick	Greater Geelong	Barwon S/W	
12	AFL Auskick - Anakie	2155 Ballan Roac	Anakie	3221	VIC	AFL Auskick	Surf Coast , Grea	Grampians, Barwon S/W	
13	AFL Auskick - Anglesea	Ellimatta Reserve	Anglesea	3230	VIC	AFL Auskick	Surf Coast	Barwon S/W	
14	AFL Auskick - Apollo Bay	Apollo Bay Recrea	Apollo Bay	3233	VIC	AFL Auskick	Colac-otway	Barwon S/W	
15	AFL Auskick - Apollo Parkways		Greensborough	3088	VIC	AFL Auskick	Nillumbik ; Bany	North and West Metropolitan Region	
16	AFL Auskick - Ararat	Richardson Oval	Ararat	3377	VIC	AFL Auskick	Ararat Rural	Grampians	
17	AFL Auskick - Ararat North	Blake St	Ararat	3377	VIC	AFL Auskick	Program		

FIGURE 7.22 Data set downloaded into spreadsheet software

Spreadsheet software also contains a feature know as a web query that allows data to be acquired from data sources. One limitation of a web query is that the data needs to be recognised by the spreadsheet software.

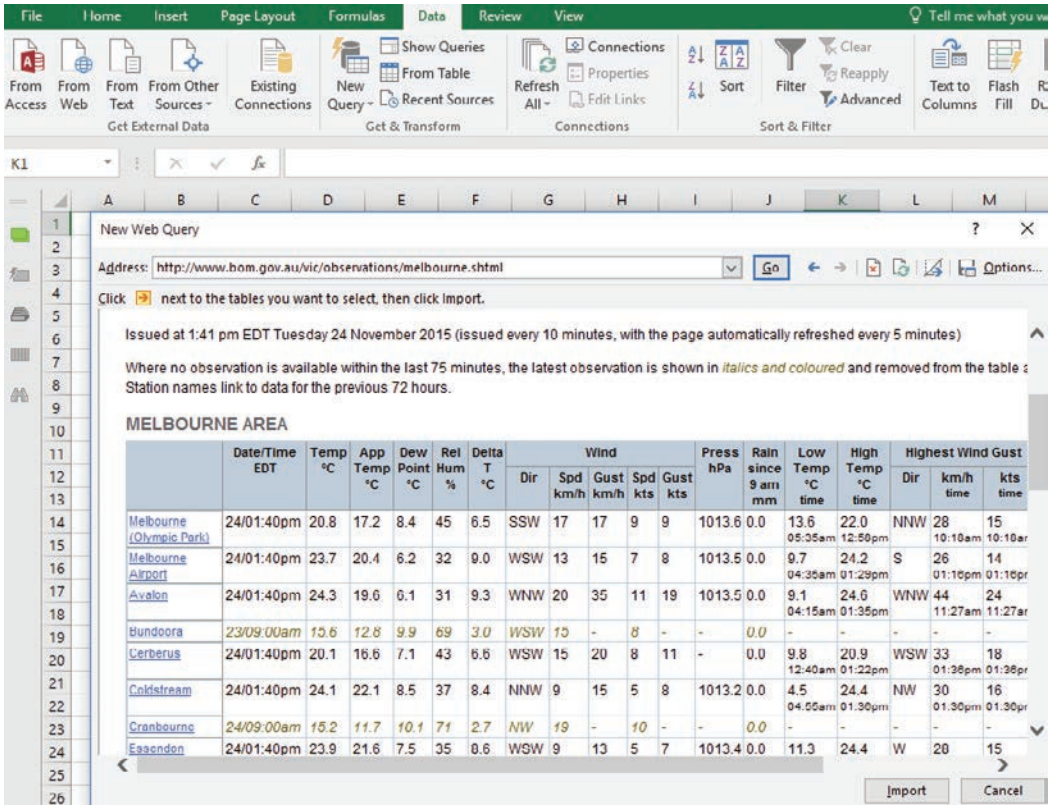


FIGURE 7.23 Example of a web query in spreadsheet software

	A	B	C	D	E	F	G	H	I	J	K	L	M
121		Date/Time	Temp	App	Dew	Rel	Delta	Wind					Press
122		EDT	°C	Temp	Point	Hum	T						hPa
123				°C	°C	%	°C						
124								Dir	Spd	Gust	Spd	Gust	
125									km/h	km/h	kts	kts	
126	Melbourne (Olympic Park)	24/01:50pm	20.6	17.5	8.9	47	6.2	SW	15	20	8	11	1014
127													
128	Melbourne Airport	24/01:50pm	23.6	18.9	5.2	30	9.3	SSW	19	22	10	12	1013
129													
130	Avalon	24/01:50pm	24.1	18.6	5.9	31	9.3	WNW	24	41	13	22	1013
131													
132	Bundoora	23/09:00am	15.6	12.8	9.9	69	3	WSW	15	-	8	-	-
133	Cerberus	24/01:50pm	20.2	15.9	6.8	41	6.8	WSW	19	22	10	12	-
134													
135	Coldstream	24/01:50pm	24.6	21	6.8	32	9.3	N	15	22	8	12	1013
136													
137	Cranbourne	24/09:00am	15.2	11.7	10.1	71	2.7	NW	19	-	10	-	-
138	Essendon Airport	24/01:50pm	23.9	19.6	7.1	34	8.8	SSW	19	20	10	11	1013
139													
140	Fawkner Beacon	24/01:50pm	-	-	-	-	-	SSE	19	20	10	11	-
141													
142	Ferny Creek	24/01:50pm	19.1	17.7	8.7	51	5.6	NW	6	7	3	4	-
143													
144	Frankston	24/01:50pm	17.2	15.9	9.3	59	4.1	WNW	6	6	3	3	-
145													
146	Geelong Racecourse	24/01:50pm	23.1	18.6	6.8	35	8.4	W	20	32	11	17	-
147													
148	Laverton	24/01:50pm	21.5	17.2	9.4	46	6.5	SE	22	28	12	15	1014
149													
150	Moorabbin Airport	24/01:50pm	20.6	16.6	9.2	48	6	SSW	20	28	11	15	1014
151													
152	Phillip Island	24/09:00am	15.7	11.4	10.6	71	2.7	WNW	24	-	13	-	-
153	Point Wilson	24/01:50pm	-	-	-	-	-	WSW	20	26	11	14	-

FIGURE 7.24 Data imported into spreadsheet software using a web query

Database software

Database software can also be used to acquire data from data sources. Using the external data function, data can be imported from a range of sources and file formats including HTML documents, Excel, text and XML files, as well as external databases.

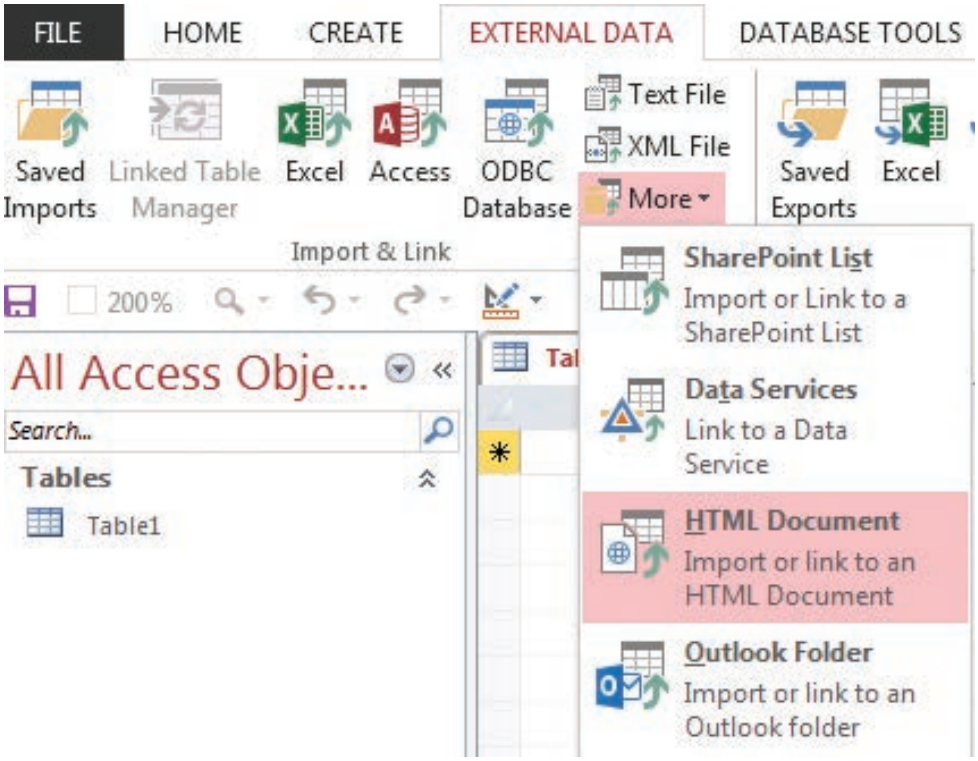


FIGURE 7.25 Options to import external data into database software

Programming languages

Some programming languages have the capability to acquire data from external sources. For example, Python can fetch data from websites. Applications can be created in Python that will fetch data from a website based on the page's HTML code.

Using the 'Inspect Element' command in a browser, you can view the HTML code of a webpage. The area that the data is located in the code can be identified and then imported into another application. Once data has been acquired, it can be imported into other applications, such as database or spreadsheet software.

Figure 7.26 shows a sample of Python code that will read all of the HTML code from Google.

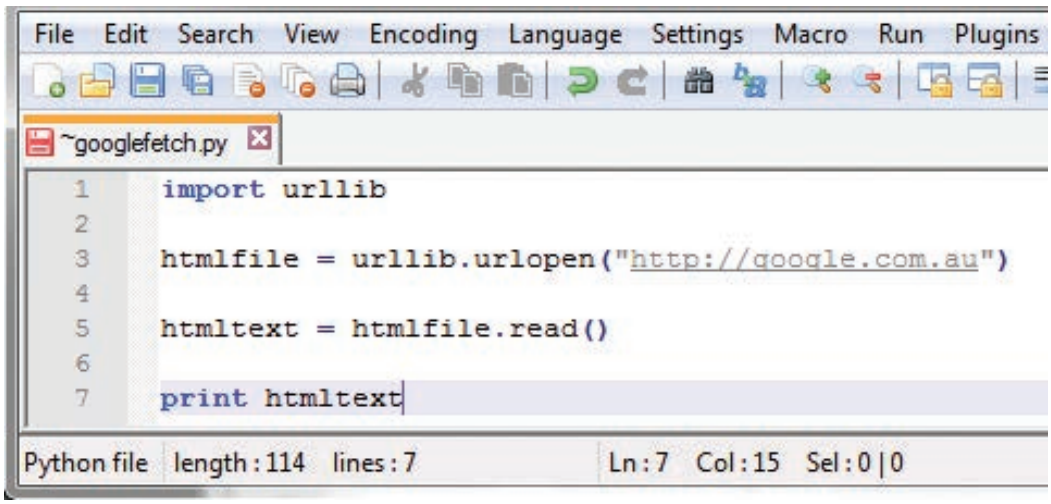


FIGURE 7.26 Sample Python code reading HTML code from Google

Displaying data

The following list shows tools and functions that can be used to display data.

Spreadsheet software

Many spreadsheet applications have a chart or graphing feature that helps manipulate the data into a visual form. Microsoft Excel has a range of different charts available.

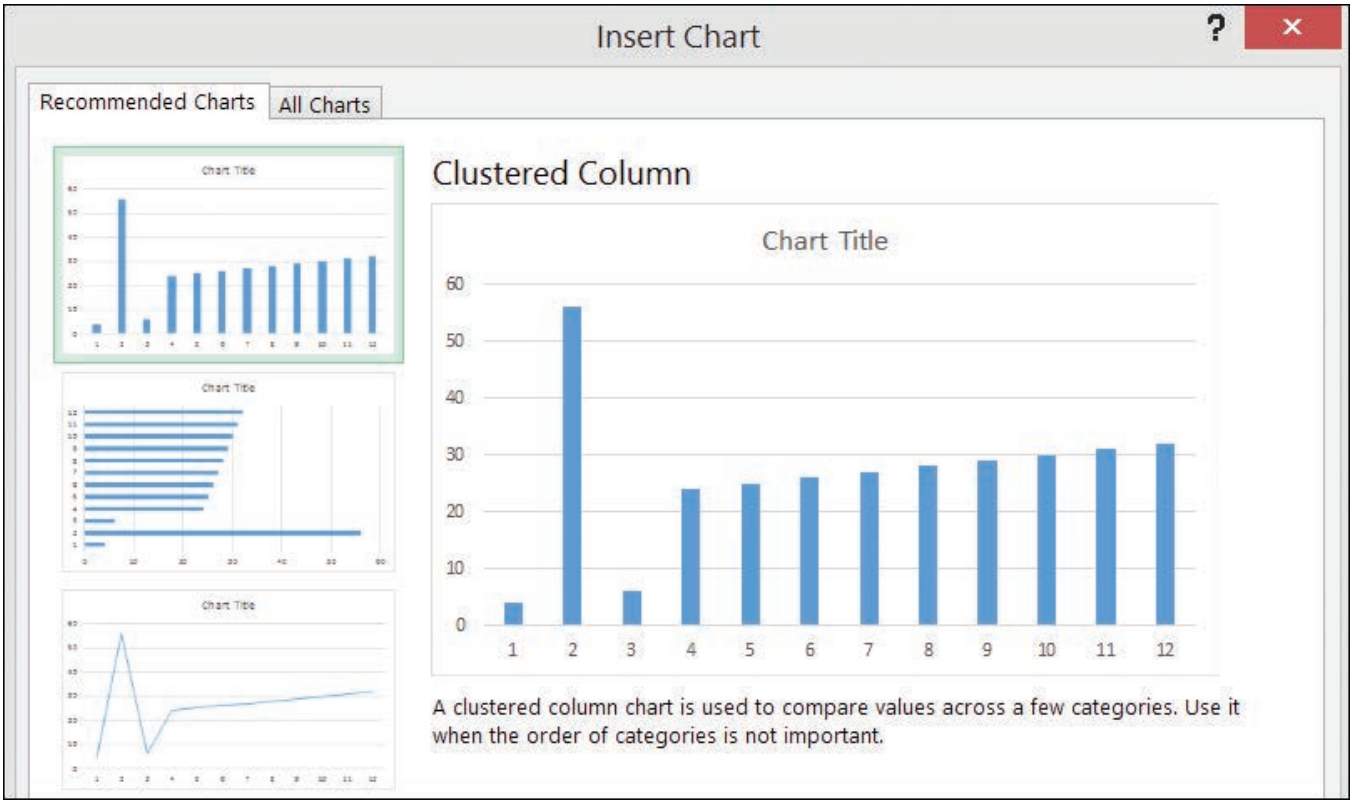


FIGURE 7.27 Spreadsheet chart function

Google

Google has a range of online cloud-based software tools that can manipulate data into a visual format. Google Sheets is spreadsheet software, Google Charts offer a range of live streaming of data from websites, and Google Motion has the functionality to develop motion charts that can be used to create time visualisations.

FIGURE 7.28 Motion chart created with Google Docs

Tableau Public

Tableau Public is free software that can allow anyone to connect to a spreadsheet or file to create interactive data visualisations for the web.

OpenHeatMap

OpenHeatMap allows users to create static or animated heat maps. Data can be saved into files including Google Docs or spreadsheets and then uploaded to the site to create the map. Heat maps represent values in a range of colours that indicate concentration, similar to a weather map.

Tableau Software

FIGURE 7.29 Tableau Public

THINK ABOUT COMPUTING 7.11

The tools mentioned are merely a sample of those available online. Research other online tools and functions available to create data visualisations.

File formats

Earlier in the chapter, files were discussed as a type of data structure. A data structure is a way to organise a collection of data items or a data set. Often the data set required to help create a data visualisation will be stored in a file. But just as there are many types and formats of data visualisations, there are many file types and formats that store the data sets themselves.

Below we will look at some common file formats used to store data used in visualisations.

XLM/XLMS file formats

A file with the **XLM** or **XLSX file** extension is an Excel Microsoft Office Open XML Format Spreadsheet file. Microsoft Office Excel is the primary software program used to open and edit XLM and XLSX files. Alternatively there are a range of free alternative spreadsheet programs that can use these files.

WMS file formats

A **WMS file** is a Web Map Service standard that involves retrieving map images over the internet from a webserver. WMS files are used in conjunction with GIS file formats to create geospatial visualisations.

FIGURE 7.30 Web Map Service (WMS) using geographic information system (GIS) file data to show industrial areas of urban development. The areas outlined with purple lines are flagged as 'industrial nodes'.

GIS file formats

A **GIS file** format contains geographical data. For example, it could contain the location of rivers and creeks in Victoria, stored as a series of longitude and latitude coordinates. For each river or creek, it could also contain attribute information, including the name and some history information. The GIS file will also contain data on how to display each creek and river in the visualisation.

CSV file format

A **comma-separated value (CSV) file** stores data in tabular format in a plain text file. The data is saved as individual fields using a comma to separate values. There are a number of advantages of saving data in a CSV format. First, a CSV is quite easy to import into (or export from) a range of software applications, including spreadsheet and database applications. Plain text files (such as notepad) also use very little storage space. Unlike some other file formats, they are saved as text, so are readable.

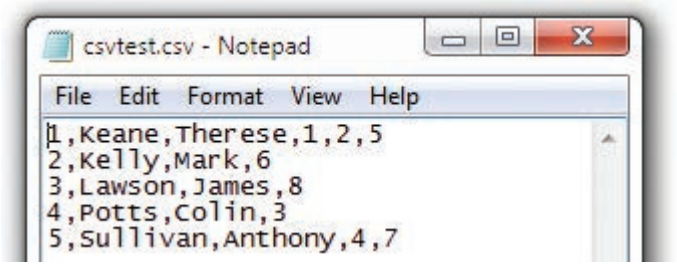


FIGURE 7.31 Comma separated value file format

API tools

An **application-programming interface (API) tool** is more a tool than a file format, but does allow the data required for a visualisation to be collected. An API tool contains the instructions and standards to allow live data to be collected from external websites and then be used in a data visualisation.

THINK ABOUT COMPUTING 7.12
There are many more file formats for storing data sets used in visualisations. Create a list of five other file formats you could use.

Evaluating data visualisations

Evaluation involves measuring how well the solution meets the information need and the needs of the client. Evaluation is used in two stages of the problem-solving methodology. First evaluation criteria are created in the design stage. Then, in the evaluation stage, two activities occur: determining an evaluation strategy and then completing an evaluation report.

The evaluation stage is not completed until after the solution is developed and been used by actual users for a period of time, often three to six months after development. This time frame allows the users to learn and become familiar with the solution, so they can make clear judgements on whether the solution is meeting their information needs and any other requirements.

Evaluation criteria

Evaluation criteria are measures that will be used to judge whether the solution meets the information needs of the client or user.

Criteria should relate to the initial solution requirements because if all the requirements are contained within the solution, then the information need should be met. As the criteria are going to be used to judge the solution, they should be framed or written so that they are quantifiable or measurable.

OpenHeatMap: World Bank Data Explorer

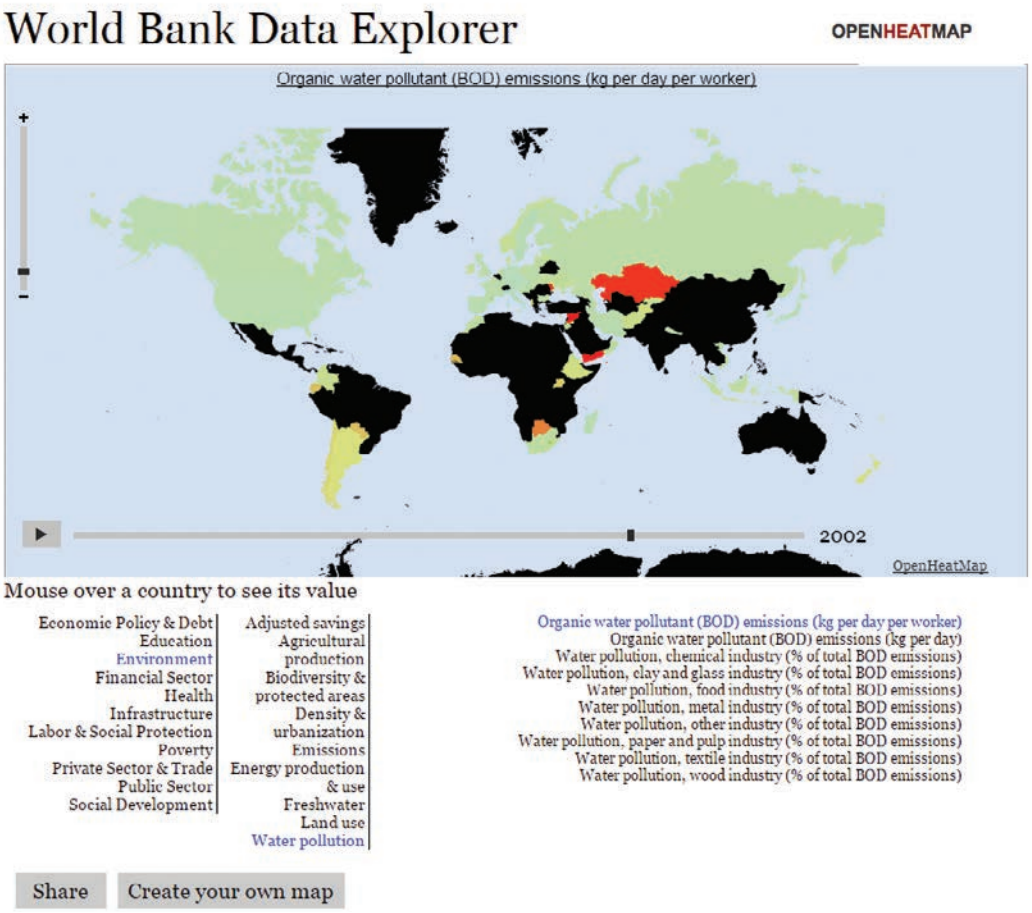


FIGURE 7.32 Dynamic heat map created with OpenHeatMap

Quantifiable or measurable criteria are measures whose satisfaction is easily determined. An example of a measurable criterion is 'the solution displays the visualisation within five seconds of loading the website'. It is relatively easy to then judge whether the solution can do this, or not.

Evaluation criteria should also cover both **effectiveness** and **efficiency**. Effectiveness criteria should relate to how well the solution works and if it provides the information needed. There are a number of characteristics related to effectiveness including accuracy, accessibility, attractiveness, communication of message, completeness, timeliness and useability. For a solution to be considered effective, it needs to have these characteristics. Based on the sample solution requirements, listed on pages 291–2, examples of evaluation criteria for effectiveness include:

- population sizes of towns are represented accurately
- visualisation contains appropriate contrast, space and balance
- data used is timely
- data for all towns in Victoria is included
- solution is easy to use.

Efficiency criteria relate to saving time, cost or effort when retrieving the information from the solution. Examples of evaluation criteria for efficiency include:

- town information can be accessed within three seconds
- zooming in and out can be completed without using the keyboard
- solution will reduce the amount of money spent on printing the data.

The evaluation criteria that have been created during the design stage are then available for the developers to refer to during the development stage to provide guidance for the project.

Evaluation strategy

The evaluation stage of the problem-solving methodology is not completed until after the solution is created and is being used by real users. As discussed, often a timeframe of three to six months after implementation is considered a reasonable time for the evaluation stage to begin. The first activity in the evaluation stage involves creating an evaluation strategy. An evaluation strategy involves deciding how each evaluation criteria can be measured. It includes creating a timeline for evaluation to take place, deciding on the data required to help judge each criterion, and looking at the way the data required will be collected and how the data can be used to evaluate each measure.

One method to display an evaluation strategy, using one criterion for effectiveness and one for efficiency from the example above, is to use a table format, as shown in Table 7.5.

TABLE 7.5 The evaluation strategy

Evaluation criterion	Data used must be less than 6 months old	Town information can be accessed within 3 seconds
Timeline	3 months after solution implementation	3 months after solution implementation
Data required	Date that the data used to create the visualisation	Time taken for a user to zoom in to a particular town and access the town information
Data collection method	Interview the staff involved in collecting the primary data about population to establish the date it was collected.	Observe a user use the solution and, using a stopwatch, time how long it takes to access the information of a particular town.
How the data can be used	Compare the date that the data was collected to the current date to determine how old the data is.	Compare the average time taken, over 10 attempts, to access the information to the benchmark (3 seconds).

A range of methods can be used to collect the data required for evaluation. We have previously discussed using interviews, survey and observations to collect data. In addition, checking download speeds, counting website hits, inspecting the solution output, reviewing error logs or timing how long it takes users to complete tasks are just some additional methods that enable evaluation criteria to be completed.

THINK ABOUT COMPUTING 7.13

Create a strategy table by selecting two other of the sample criteria listed on page 290. Complete the evaluation strategy for each.

Evaluation report

Finally, after the evaluation strategy has been completed, the evaluation report can be written. The evaluation report involves stating if the solution is meeting the information need and the needs of the user. To provide evidence for the final conclusion, each evaluation criterion must be assessed to identify if it has been achieved. If each of the evaluation criteria has been met, then the solution can be considered a success.

TABLE 7.6 Assessing each criterion to assist in preparing an evaluation report

Evaluation criterion	Data used must be less than 6 months old	Town information can be accessed within 3 seconds
Timeline	3 months after solution implementation	3 months after solution implementation
Data required	Date that the data used to create the visualisation	Time taken for a user to zoom in to a particular town and access the town information
Data collection method	Interview the staff involved in collecting the primary data about population to establish the date it was collected.	Observe a user use the solution and, using a stopwatch, time how long it takes to access the information of a particular town.
How the data can be used	Compare the date that the data was collected to the current date to determine how old the data is.	Compare the average time taken, over 10 attempts, to access the information to the benchmark (3 seconds).
Data	4 months	3.6 seconds
Achieved	Yes	No

CHAPTER SUMMARY

ESSENTIAL TERMS

accuracy correctness, without mistakes

application-programming interface (API) a tool that contains the instructions and standards to allow live data to be collected from external websites and then be used in a data visualisation

array a collection of data items generally of the same data type

authenticity in relation to data, data that is genuine, original and considered trustworthy

Boolean data type a logical data type; Boolean data can hold only two possible values, usually true or false

character a single letter, number or symbol that is usually stored in 2 bytes of memory; a data type

chart a method of displaying data as symbols

comma-separated value (CSV) file a file type that stores data in tabular format in a plain text file; the data is saved as individual fields using a comma to separate values

data types particular forms that an item of data can take, including numeric, character and Boolean

data visualisation the presentation of data in a pictorial or graphical format

floating point number a number with a fractional or decimal component

flow visualisations representing data that illustrates the flow pattern of an object or item

formats ways in which data and information can be presented

functional requirements those directly related to what the solution will do

geospatial visualisation overlay data related to a geographical area over mapping technology; Google Maps is an example of a geospatial visualisation

GIS file a file format that contains geographical data

hierarchy visualisations show the relationships and structure between data items

information need when particular information is required, yet is not being currently supplied; causes of an information need may include a current problem, an identified need or an opportunity

integer a number without a fractional or decimal component

map-based visualisations geographical data displayed in a visual format

matrix visualisations visualisations used to show the compensation of individual items in the sample size

network visualisations visualisations that show the relationship between different data items and data sets

non-functional requirements other requirements that the user or client would like the solution to have but they do not affect what the solution does

non-technical constraints limitations relating to areas other than hardware and software: social, legal and useability

record a collection of data items of different data types

relevance data that produces useful information required by the information need; if the data is not useful, it is not relevant

sample size the number of completed responses that you collect from interviews or surveys; the more responses completed the more likely the data being collected is accurate

solution constraints factors that may affect or restrict the requirements included in the solution

solution requirements what the client wants from the solution; can be broken down into functional and non-functional requirements

string data that consists of a string of characters

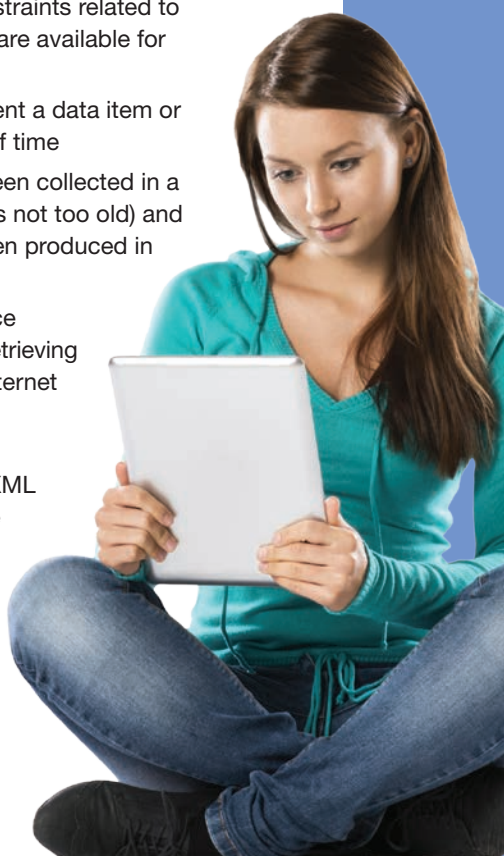
technical constraints constraints related to the hardware and software available for the project

time visualisations represent a data item or data set over a period of time

timeliness data that has been collected in a reasonable timeframe (is not too old) and information that has been produced in time to be useful

WMS file a web map service standard that involves retrieving map images over the internet from a webserver

XLM or XLSX file an Excel Microsoft Office Open XML Format Spreadsheet file



CHAPTER SUMMARY

IMPORTANT FACTS

1 Government departments collect and store vast amounts of data sets.
2 The Bureau of Meteorology is Australia's national weather, climate and water agency.
3 The Victorian government data directory contains a range of data sets from Victorian government departments.
4 The Australian Bureau of Statistics (ABS) is the statistical agency of the federal government.
5 Data.gov.au is similar to the Victorian government data directory in that it provides publically available data sets that have been collected by federal government departments.
6 A file data structure can hold numerous data items, arrays or records. A file is saved separately from the software program that utilises the file.
7 Infographics generally are used to communicate a particular message.
8 The first stage of the problem-solving methodology is the analysis stage. The analysis stage is considered with 'what' questions. The analysis stage is made up of three activities: determining solution requirements, solution constraints and the scope of the solution.
9 Economic factors include time and money.
10 Scope of solution helps to identify the boundaries or parameters of the solution and outlines what requirements will or will not be included.
11 The design stage is all about how the solution will be achieved.
12 The solution design involves planning how the solution will appear and function.
13 A layout diagram can be used to plan how the visualisation will appear.
14 Storyboards can be used to show how the data-visualisation animation might work.
15 An IPO chart is a design tool that helps to plan how a solution will function.
16 Flowcharts can be used to show the procedure that users need to complete to create a data visualisation.
17 Conventions are general rules that are followed when using a particular format.
18 Common conventions include clear title, axis labelled, key or legend used, name of the author and the source of the data, unit of measurements shown and matching colours.
19 There are a range of software tools and functions that can be used to create data visualisations.
20 Many spreadsheet applications have a chart or graphing feature that helps manipulate the data into a visual form.
21 Google has a range of online cloud-based software tools that can manipulate data into a visual format.
22 Tableau Public is free software that can allow anyone to connect to a spreadsheet or file to create interactive data visualisations for the web.
23 OpenHeatMap allows users to create static or animated heat maps.
24 A file is a type of data structure that can store and organise a data set. Files operate independently of the software application that uses the file, so after the application is closed, the file will still exist. There are many file types and formats.
25 A file with the XLM or XLSX file extension is an Excel Microsoft Office Open XML Format Spreadsheet file.
26 An application-programming interface (API) tool is more a tool than a file format but does allow the data required for a visualisation to be collected.
27 Evaluation involves measuring how well the solution meets the information need and the needs of the client.
28 In the design stage of the problem-solving methodology, evaluation criteria are created.
29 Evaluation criteria should also cover both effectiveness and efficiencies. Effectiveness criteria should relate to how well the solution works and if it provides the information needed.
30 Efficiency criteria relates to saving time, cost or effort in retrieving the information from the solution.
31 The evaluation stage of the problem-solving methodology is not completed until after the solution is created and is being used by real users.
32 An evaluation strategy is about deciding how each evaluation criteria can be measured.
33 Interviews, surveys, observations, checking download speeds, counting the number of website hits, inspecting the solution output, reviewing error logs and timing users are all methods used to collect evaluation data.
34 The evaluation report involves stating if the solution is meeting the information need and the needs of the user. To provide evidence for the final conclusion, each evaluation criterion must be assessed to identify if it has been achieved.

TEST YOUR KNOWLEDGE

INFORMATION NEEDS AND DATA VISUALISATIONS

Review quiz

1 Describe an information need.
2 List three reasons why an information need may occur.
3 Explain when an information need may result in a data visualisation being developed.

SOURCES OF AUTHENTIC DATA

4 Explain the difference between primary and secondary sources of data.
5 List three sources of primary data.
6 Explain when a survey would be better than an interview to collect data.
7 List three sources of secondary data.
8 Identify an issue related to using data from secondary sources.
9 Explain why government departments tend to be the source of numerous data sets.
10 Explain the role of the Bureau of Meteorology.
11 Identify how the Victorian government data directory differs from Data.gov.au.
12 Explain how the role of the Australian Bureau of Statistics differs from that of Data.gov.au.

INTEGRITY OF DATA

13 Define 'integrity of data'.
14 List four measures used to judge the integrity of data.
15 Outline a measure that can be used to improve the chances of data being accurate when collected.
16 Timeliness does not mean as quickly as possible. Describe the concept of timeliness.
17 List three attributes of authentic data.
18 For creating a visualisation, what would be considered relevant data?

DATA TYPES AND DATA STRUCTURES

19 Define 'data type'.
20 List four data types.
21 Explain the relationship between a character data type and a text (string) data type.
22 Explain how an integer differs from a floating point.
23 If collecting data for the question 'Are you 18 or over?', which data type would be the most efficient in storing the data? Give reasons for your answer.
24 Explain how an array is different from a record.
25 Give two reasons why files are an effective method of storing data sets.

TYPES AND PURPOSES OF DATA VISUALISATION

26 Define 'data visualisation'.
27 List three types of visualisations that can fit into the chart category.
28 Explain the concept of geospatial visualisations.
29 Discuss one advantage of geospatial visualisations.
30 Describe a situation when you would use a network visualisation.
31 Explain how timeline data differs from time series data.
32 List three examples of hierarchy visualisations.
33 Explain the purpose of a flow visualisation. Give an example to support your answer.
34 Discuss how a matrix visualisation is often laid out.

ANALYSIS STAGE

35 List the three activities involved in the analysis stage of the problem-solving methodology.
36 Define 'solution requirements'.
37 Explain the difference between functional requirements and non-functional requirements.
38 Define 'functional constraint' and 'non-functional constraint'.
39 Explain how constraints can affect the solution.
40 State the purpose of defining the scope of the solution.

DESIGN TOOLS

41 Explain the purpose of the design stage.
42 List the two activities involved in the design stage.
43 Outline the purpose of a design tool.
44 Explain the difference between appearance design tools and functionality design tools.
45 List two design tools for appearance.
46 Discuss the advantage of creating a number of layout diagrams representing the data in a different format.
47 Explain the purpose of an IPO chart.
48 Outline a situation when using a flowchart would be appropriate.

FORMATS AND CONVENTIONS

49 Define 'format'.
50 Define 'convention'.
51 List four conventions usually followed in data visualisations.

SOFTWARE TOOLS AND FUNCTIONS

52 List three software tools used to create data visualisations.
53 Explain the purpose of a motion chart.
54 Explain one similarity between Tableau Public and OpenHeatMap.

FILE FORMATS

55 Explain the relationship between files and software applications.
56 List three types of files commonly used to store data sets.
57 Explain the relationship between WMP files and GIS files.
58 Describe how a CSV file is structured.
59 Outline the purpose of an API tool.

EVALUATING VISUALISATIONS

60 Define 'evaluation'.
61 Identify when an evaluation takes place. Explain why an evaluation takes place at this point.
62 List the two activities involved in the evaluation stage.
63 Identify an activity completed in another stage of the problem-solving methodology that provides input into the evaluation stage.
64 Explain the purpose of the evaluation strategy.
65 List three methods that can be used to collect evaluation data.
66 Discuss the purpose of an evaluation report.
67 Explain the link between the evaluation criteria and whether a solution can be deemed a success.

APPLY YOUR KNOWLEDGE

Using the following data sets, create visualisations that will provide useful information.

1 Analyse the data to determine, in any three-month period (for example, Feb to Apr) which city would be the best to visit for warmer weather.

TABLE 7.7 Sydney versus Melbourne weather

City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Melbourne	25.9	25.8	23.9	20.3	16.7	14.1	13.5	15.0	17.2	19.7	22.0	24.2
Sydney	25.9	25.8	24.7	22.4	19.4	16.9	16.3	17.8	20.0	22.1	23.6	25.2

2 **a** Calculate each state's proportion of Australia's population and display as a data visualisation. All states have twelve senators in the Australian Senate, while the Australian Capital Territory and the Northern Territory have two.

b Calculate the number of people in each jurisdiction for each senator. Create a data visualisation of the results.

TABLE 7.8 Australian population by state and territory

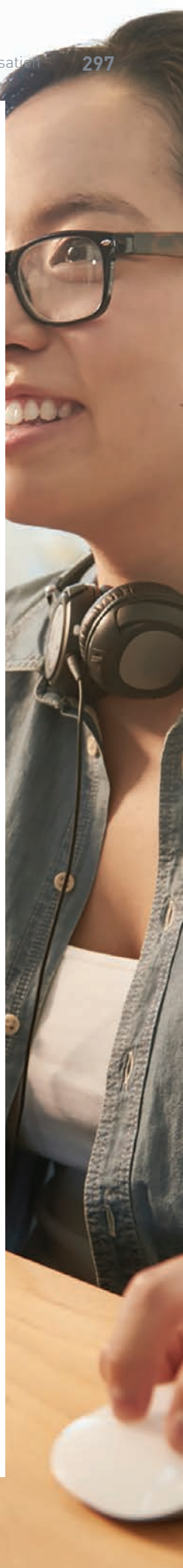
State	Population
New South Wales	7 272 800
Victoria	5 603 100
Queensland	4 516 361
Western Australia	2 296 411
South Australia	1 644 642
Tasmania	507 626
Australian Capital Territory	358 894
Northern Territory	229 675

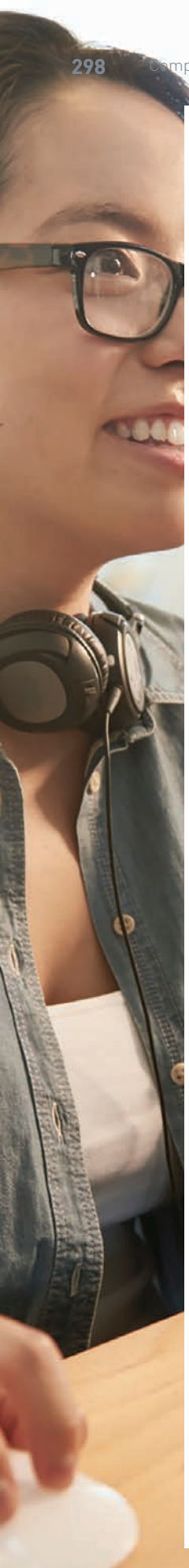
3 Create a visualisation showing the total number of students at the school and the proportion of boys to girls over time.

TABLE 7.9 Student enrolments at the Melbourne School

Year	1961	1971	1981	1991	2001	2011
Boys	81	450	600	750	750	900
Girls	0	0	0	0	150	400
Total	81	450	600	750	900	1300

These figures are estimates.





4 Create a visualisation to display the number of members of each AFL club from each state, assuming that all members live in the state where their team resides.

Collingwood Magpies	77 405	Port Adelaide Power	40 510
Hawthorn Hawks	62 393	Sydney Swans	35 612
Richmond Tigers	60 053	North Melbourne Kangaroos	34 511
West Coast Eagles	57 830	St Kilda Saints	32 562
Essendon Bombers	54 903	Melbourne Demons	32 847
Carlton Blues	49 224	Western Bulldogs	29 641
Adelaide Crows	45 000	Brisbane Lions	23 760
Fremantle Dockers	43 638	Greater Western Sydney Giants	12 631
Geelong Cats	41 935	Gold Coast Suns	12 350

PREPARING FOR UNIT 2 OUTCOME 2

Produce a data visualisation that meets the need of a user. This will involve using a complex data set and accessing software or online tools that will enable you to convert this data into a more meaningful format

Apply the problem-solving methodology and use appropriate software tools to create a data visualisation to meet user needs.

OUTCOME MILESTONES

1. Familiarise yourself with the design brief and analyse the information problem that exists.
2. Determine the type of data visualisation needed to solve the information problem.
3. Select appropriate sources of data and identify relevant data.
4. Determine the suitability of different data types and structures for creating visualisations.
5. Select types of visualisations that are appropriate to the data.
6. Select and apply appropriate tools to plan the design of the visualisations.
7. Apply software functions to locate and acquire data to input and manipulate.
8. Use appropriate software tools, and select and apply a range of suitable functions.

STEPS TO FOLLOW

Create the solution using the four stages of the PSM: Analysis, design, development and evaluation.

Analysis stage

1. Determine the solution requirements.
2. Determine the solution constraints.

Design stage

1. Plan the solution: Use design tools for appearance.
2. Plan the solution: Use design tools for functionality.
3. Establish evaluation criteria.
4. Extract relevant data from data sources.
5. Manipulate data for use.

Development stage

Create a data visualisation using appropriate software types and functions.

Evaluation stage

1. Create criteria to evaluate if the data visualisation meets users' needs.
2. Select appropriate techniques to determine if the data visualisation meets users' needs.