data analytics

VCE UNITS 3 & 4

5E

GARY BASS

NATALIE HEATH

THERESE KEANE

ANTHONY SULLIVAN

SIGN IN & SUCCEED
NELSONNET ACCESS CODE INSIDE!

Data Analytics VCE Units 3&4
5th Edition
Gary Bass
Natalie Heath
Therese Keane
Anthony Sullivan
Mark Kelly

Senior publisher: Eleanor Gregory
Editor: Scott Vandervalk
Proofreader: Nadine Anderson
Indexer: Bruce Gillespie
Visual designer: James Steer
Cover design: Chris Starr, MakeWork
Text design: Leigh Ashforth, Watershed Art & Design
Permissions researcher: Lyahna Spencer
Production controller: Karen Young
Typeset by: DiacriTech

Acknowledgements
Extracts from the VCE Applied Computing Study Design (2020–2023), are reproduced by permission, © VCAA. VCE is a registered trademark of the VCAA. The VCAA does not endorse or make any warranties regarding this study resource. Current VCE Study Designs, past VCE exams and related content can be accessed directly at www.vcaa.vic.edu.au

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email
aust.permissions@cengage.com

ISBN 978 0 17 044087 5

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit **cengage.com.au**

Printed in Singapore by 1010 Printing International Limited.
1 2 3 4 5 6 7 23 22 21 20 19

Contents

Preface v
About the authors vi
How to use this book vii
Outcomes ix
Problem-solving methodology xiii
Key concepts xvi

Unit 3

Introduction 1

Chapter 1 Data and presentation 2

What is data? 3
Referencing data sources 14
Data types 16
Data structures 19
Data integrity 20
Data validation 24
Data visualisation 25
Designing solutions 37
Design principles 37
Formats and conventions 39
Design tools for data visualisation 46

Chapter 2 Data manipulation and presentation 57

Databases 58
Creating a database structure 59
Spreadsheet tools 84
Data visualisation 89
Testing 96
Review the Outcome's steps 106
Preparing for Unit 3, Outcome 1 112

Chapter 3 Project management and data analysis 113

What is data? 114
Project management 114
Why must we begin with a research question? 124
Primary and secondary data 128
Quantitative and qualitative data 129
Data types and data structures 136
Referencing data sources 137
Data integrity 139
Data security 152
Next steps 158

Chapter 4 Drawing conclusions 165

Continuing Unit 3, Outcome 2 166
Solution specifications 166
Design principles 171
Generating design ideas 176
Design tools 187
Types of infographics and data visualisations 190
Preparing for Unit 3, Outcome 2 204

9780170440875

Unit 4

Introduction **207**

Chapter 5 Development and evaluation 208

Data visualisations 209
Procedures and techniques for managing files 214
Functional capabilities of software 220
An effective infographic or data visualisation 221
Manipulating data 234
Formats and conventions 237
Manipulating data with software 238
Verification and validation 243
Testing 244
Evaluating your solution 248
Documenting the progress of projects 250
Assessing your project plan 254
Next steps 255
Preparing for Unit 4, Outcome 1 261

Chapter 6 Information management 262

Networks 263
Threats to data and information 272
Physical and software security controls 275
Managing data on a network 284
Network attached storage and cloud computing 289

Chapter 7 Cyber security measures 297

The importance of data and information to organisations 298
Information management strategies 301
Data security 302
The importance of diminished data integrity in information systems 303
Key legislation relating to data and information 304
Ethics and security practices 314
Resolving legal and ethical tensions 317
Reasons to prepare for disaster 318
Consequences of security failure 328
Evaluating information management strategies 331
Preparing for Unit 4, Outcome 2 338

Index **339**

9780170440875

Preface

This fifth edition of *Data Analytics VCE Units 3 & 4* incorporates the changes to the VCAA VCE Applied Computing Study Design that took effect from 2020.

This textbook looks at how individuals and organisations use, and can be affected by, information systems in their daily lives.

We believe that teachers and students require a text that focuses on the **Areas of Study** specified in the **Study Design** and which presents information in a sequence that allows easy transition from theory into practical assessment tasks. We have, therefore, written this textbook so that a class can begin at Chapter 1 and work their way systematically through to the end. Students will encounter material relating to the **key knowledge** dot points for each **Outcome** before they reach the special section that describes the Outcome. The Study Design outlines **key skills** that indicate how the knowledge can be applied to produce a solution to an information problem. These Outcome preparation sections occur regularly throughout the textbook and flag an appropriate point in the student's development for each Outcome to be completed. The authors have covered all key knowledge for the Outcomes from the Data Analytics VCE Units 3 & 4 course.

Our approach has been to focus on the key knowledge required for each school-assessed Outcome, and to ensure that students are well prepared for these; however, there is considerable duplication in the Study Design relating to the knowledge required for many of the Outcomes. We have found that, with an Outcomes approach, we are covering material sometimes several times. For example, knowledge of a problem-solving methodology is listed as key knowledge for many different Outcomes. In these cases, we have tried to provide a general coverage in the first instance, and specifically apply the concept to a situation relevant to the related Outcome on subsequent encounters.

The authors assume that teachers will develop the required key skills with their students within the context of the key knowledge addressed in this textbook and the resources available to them.

We have incorporated a margin column in the text to provide additional information and reinforce key concepts. This margin column also includes activities that relate to the topics covered in the text and considers issues relevant to information systems usage.

Outcome features appear at several points in the book, indicating the nature of the tasks that students undertake in the completion of the school-assessed Outcome. We have listed the steps required to complete the Outcome, together with advice and suggestions for approaching the task. We have also described the output and support material needed for submission. You will also find sample tasks and further advice relating to the Outcomes are available at **https://nelsonnet.com.au**.

The chapters are organised to present the optimum amount of information in the most effective manner. The book is presented in concise, clearly identified sections to guide students through the text. Each chapter is organised into the sections described on pages **vii–viii**.

9780170440875

About the authors

Gary Bass teaches VCE Applied Computing at Year 11 and Year 12 in an online course environment at Virtual School Victoria. Previously, he has taught VCE Physics, as well as developing and delivering middle school ICT courses. Gary has presented at DLTV DigiCON and the annual IT teachers conference on many topics including Pop-up Makerspace; Big Data requires huge analysis – data visualisation; AR + VR = Mixed reality; and Marshall McLuhan-Medium is the message. Gary was selected as an Apple Distinguished Educator (ADE) in 2002 and 2011. In 2016, he was presented with DLTV's IT Leader of the Year award.

Natalie Heath is the eLearning/ICT Leading Teacher at Eltham High School. She has been an IT specialist teacher for nearly 20 years, teaching at all secondary levels including VCE Informatics, IT Applications, Information Processing and Management, and Software Development. She has extensive experience around VCE, having assessed examinations in various subjects for more than two decades. Natalie has also developed many resources for VCE Computing subjects over the years, including trial examinations. She has presented at teacher professional learning conferences as an expert in Unit 3 and 4 subjects and in 2018 was presented with the DLTV's Maggie Iaquinto VCE Computing Educator of the Year award for her services to the VCE Computing teaching community and resource development.

Associate Professor Therese Keane is Deputy Chair of the Department of Education at Swinburne University and has worked in a variety of school settings where she has taught IT in K–12 education as the Director of ICT. Her passion and achievements in ICT in the education and robotics space have been acknowledged by her peers in her receiving numerous national and state awards. Therese has presented seminars and workshops for teachers involved in the teaching of IT. She has written several textbooks in all units of VCE Information Technology. Therese's research interests include the use of technology in education, gender inequalities in STEM-based subjects, robotics in education and computers in schools for teaching and learning purposes. Therese is involved with the FIRST LEGO League as the Championship Tournament Director for Victoria and she is a lead mentor for the RoboCats – a female school student only robotics team that participates in the FIRST Robotic Competition.

Anthony Sullivan is a Curriculum and Learning Specialist at Monash College where he is responsible for creating assessment and learning materials for accounting and computing subjects as part of the Monash University Foundation Year program. Anthony has more than 25 years experience teaching business and computing subjects. He has taught in both government and non-government settings in Australia and taught computing and information technology courses in schools in Asia and the United Kingdom. Anthony has also been a VCE Examination Assessor, a member of the committee that reviewed and wrote the previous Study Design for VCE Computing, and has written a range of commercial resources related to VCE Computing. He has presented at conferences and professional development events and student examination preparation sessions.

9780170440875

How to use this book

KEY KNOWLEDGE

The key knowledge from the VCE Applied Computing Study Design that you will cover in each chapter is listed on the first page of each chapter. The list includes key knowledge specified in the Outcome related to the chapter.

FOR THE STUDENT

Each chapter's opening page includes an overview of that chapter's contents so that you are aware of the material you will encounter.

FOR THE TEACHER

This section is for your teacher and outlines how the chapter fits into the overall study of Data Analytics, and outlines how the material relates to the completion of Outcomes.

CHAPTERS

The major learning material that you will encounter in the chapter is presented as text, photographs and illustrations. The text describes in detail the theory associated with the stated Outcomes of the VCE Applied Computing Study Design in easy-to-understand language. The photographs show hardware, software and other objects that have been described in the text. Illustrations are used to demonstrate concepts that are more easily explained in this manner.

Throughout the chapter, glossary terms are highlighted in bold and you can find their definitions at the end of each chapter, in **Essential terms**.

The **School-assessed Task Tracker** at the bottom of every odd-numbered page provides you with a visual reminder to help you track your progress in the School-assessed Task (SAT), which is derived from Unit 3, Outcome 2 and Unit 4, Outcome 1, so that you can complete all required stages on time.

MARGIN COLUMN

The margin column contains further explanations that support the main text, weblink icons, additional material outside the Study Design and cross-references to material covered elsewhere in the textbook. Issues relevant to Data Analytics that you can discuss with your classmates are also included in the form of 'Think about Data Analytics' boxes.

THINK ABOUT DATA ANALYTICS 3.1

Project management tools are useful to find the perfect number of people needed on a task so it is finished as quickly as possible without anyone being idle. Use software to develop a Gantt chart to plan the baking of a cake. Assume you can use as many cooks as you want.

CHAPTER SUMMARY

The chapter summary at the end of each chapter is divided into two main parts to help you review each chapter.

Essential terms are the glossary terms that have been highlighted throughout the chapter.

The **Important facts** section is a list of summaries, ideas, processes and statements relevant to the chapter, in the order in which they occur in the chapter.

9780170440875

TEST YOUR KNOWLEDGE

Short-answer questions will help you to review the chapter material. The questions are grouped and identified with a section of the text to allow your teacher to direct appropriate questions based on the material covered in class. Teachers will be able to access answers to these questions at **https://nelsonnet.com.au**.

APPLY YOUR KNOWLEDGE

Each chapter concludes with a set of questions requiring you to demonstrate that you can apply the theory from the chapter to more complex questions. The style of questions reflects what you can expect in the end-of-year examination. Teachers will be able to access suggested responses to these questions at **https://nelsonnet.com.au**.

PREPARING FOR THE OUTCOMES

This section appears at points in the course where it is appropriate for you to complete an Outcome task. The information provided describes what you need to do in the Outcome, the suggested steps to be followed in the completion of the task and the material that needs to be submitted for assessment.

NELSONNET

The NelsonNet student website contains:

- multiple-choice quizzes for each chapter, mirroring the VCAA Unit 3 & 4 exam
- additional material such as spreadsheets and infographics.

An open-access weblink page is also provided for all weblinks that appear in the margins throughout the textbook. This is accessible at **https://nelsonnet.com.au**.

The NelsonNet teacher website is accessible only to teachers and it contains:

- answers for the **Test your knowledge** and **Apply your knowledge** questions in the book
- sample School-assessed Coursework (SAC)
- chapter tests
- practice exam.

Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

9780170440875

Outcomes

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 3 **Area of Study 1** **Outcome 1**	**Data analytics** On completion of this unit the student should be able to respond to teacher-provided solution requirements and designs to extract data from large repositories, manipulate and cleanse data and apply a range of functions to develop software solutions to present findings.	
Data and information	• techniques for efficient and effective data collection, including methods to collect census, Geographic Information System (GIS) data, sensor, social media and weather	pp. 6–10
	• factors influencing the integrity of data, including accuracy, authenticity, correctness, reasonableness, relevance and timeliness	pp. 20–23
	• sources of, and methods and techniques for, acquiring authentic data stored in large repositories	pp. 10–13
	• methods for referencing primary and secondary sources, including American Psychological Association (APA) referencing system	pp. 14–16
	• characteristics of data types	pp. 16–18
Approaches to problem solving	• methods for documenting a problem, need or opportunity	p. 37
	• methods for determining solution requirements, constraints and scope	p. 37
	• naming conventions to support efficient use of databases, spreadsheets and data visualisations	p. 67
	• a methodology for creating a database structure: identifying entities, defining tables and fields to represent entities; defining relationships by identifying primary key fields and foreign key fields; defining data types and field sizes, normalisation to third normal form	pp. 58–66
	• design tools for representing databases, spreadsheets and data visualisations, including data dictionaries, tables, charts, input forms, queries and reports	pp. 46–9, 84
	• design principles that influence the functionality and appearance of databases, spreadsheets and data visualisations	pp. 37–9, 84–93
	• functions and techniques to retrieve required information through querying data sets, including searching, sorting and filtering to identify relationships and patterns	pp. 74–5
	• software functions, techniques and procedures to efficiently and effectively validate, manipulate and cleanse data including files, and applying formats and conventions	pp. 80–2, 88
	• types and purposes of data visualisations	pp. 25–31
	• formats and conventions applied to data visualisations to improve their effectiveness for intended users, including clarity of message	pp. 39–45
	• methods and techniques for testing databases, spreadsheets and data visualisations	pp. 96–106
Interactions and impact	• reasons why organisations acquire data	p. 4
Key skills	• interpret solution requirements and designs to develop data visualisations	pp. 37, 167–71
	• identify, select and extract relevant data from large repositories	pp. 4, 9–13
	• use a standard referencing system to acknowledge intellectual property	p. 14
	• organise, manipulate and cleanse data using database and spreadsheet software	pp. 74–5, 80–8
	• select, justify and apply functions, formats and conventions to create effective data visualisations	pp. 39–45
	• develop and apply suitable validation and testing techniques to software tools used	pp. 24–5

9780170440875

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 3 **Area of Study 2** **Outcome 2**	Data analytics: Analysis and design On completion of this unit the student should be able to propose a research question, formulate a project plan, collect and analyse data, generate alternative design ideas and represent the preferred design for creating infographics or dynamic data visualisations.	
Digital systems	• roles, functions and characteristics of digital system components	p. 148–52
	• physical and software security controls used by organisations for protecting stored and communicated data	p. 152–8
Data and information	• primary and secondary data sources and methods of collecting data, including interviews, observation, querying of data stored in large repositories and surveys	p. 128
	• techniques for searching, browsing and downloading data sets	p. 126
	• suitability of quantitative and qualitative data for manipulation	p. 129–30
	• characteristics of data types and data structures relevant to selected software tools	p. 136–7
	• methods for referencing secondary sources, including the APA referencing system	p. 137–9
	• criteria to check the integrity of data, including accuracy, authenticity, correctness, reasonableness, relevance and timeliness	p. 139–47
	• techniques for coding qualitative data to support manipulation	p. 130–6
Approaches to problem solving	• features of a research question, including a statement identifying the research question as an information problem	p. 124–6
	• functional and non-functional requirements, including data to support the research question, constraints and scope	p. 167–71
	• types and purposes of infographics and dynamic data visualisations	p. 190–8
	• design principles that influence the appearance of infographics and the functionality and appearance of dynamic data visualisations	p. 171–6
	• design tools for representing the appearance and functionality of infographics and dynamic data visualisations, including data manipulation and validation, where appropriate	p. 187–9
	• techniques for generating alternative design ideas	p. 176–85
	• criteria for evaluating alternative design ideas and the efficiency and effectiveness of infographics or dynamic data visualisations	p. 185–7
	• features of project management using Gantt charts, including the identification and sequencing of tasks, time allocation, dependencies, milestones and the critical path	p. 114–23
Interactions and impact	• key legal requirements for the storage and communication of data and information, including human rights requirements, intellectual property and privacy	p. 307–14
Key skills	• frame a research question	p. 124–6
	• analyse and document requirements, constraints and scope of infographics or dynamic data visualisations	pp. 167–171
	• apply techniques for searching, downloading, browsing and referencing data sets	p. 126
	• select and apply design tools to represent the functionality and appearance of infographics or dynamic data visualisations	p. 187–9
	• generate alternative design ideas	p. 176–85

9780170440875

OUTCOME	KEY KNOWLEDGE	LOCATION
	• develop evaluation criteria to select and justify preferred designs	p. 185–7
	• produce detailed designs using appropriate design methods and techniques	p. 187–98
	• propose and apply appropriate methods to secure stored data	p. 152–8
	• create, monitor and modify project plans using software	p. 114–23
Unit 4 **Area of Study 1** **Outcome 1**	**Data analytics: Development and evaluation** On completion of this unit the student should be able to develop and evaluate infographics or dynamic data visualisations that present findings in response to a research question, and assess the effectiveness of the project plan in monitoring progress.	
Digital systems	• procedures and techniques for handling and managing files, including archiving, backing up, disposing of files and security	pp. 214–20
	• the functional capabilities of software to create infographics and dynamic data visualisations	p. 220
Approaches to problem solving	• characteristics of information for educating targeted audiences, including age appropriateness, commonality of language, culture inclusiveness and gender	p. 221–9
	• characteristics of efficient and effective infographics and dynamic data visualisations	p. 221–34
	• functions, techniques and procedures for efficiently and effectively manipulating data using software tools	pp. 234, 238
	• techniques for creating infographics and dynamic data visualisations	pp. 238–42
	• techniques for validating and verifying data	p. 243
	• techniques for testing that solutions perform as intended	pp. 244–8
	• techniques for recording the progress of projects, including adjustments to tasks and timeframes, annotations and logs	pp. 250–3
	• strategies for evaluating the effectiveness of infographics and dynamic data visualisations solutions and assessing project plans	pp. 248–50
Key skills	• monitor, modify and annotate the project plan as necessary	p. 251
	• propose and implement procedures for managing files	pp. 214–20
	• select and apply software functions, conventions, formats, methods and techniques to develop infographics or dynamic data visualisations	pp. 237–8
	• select and apply data validation and testing techniques, making any necessary modifications	pp. 243–8
	• apply evaluation criteria to evaluate the efficiency and effectiveness of infographics or dynamic data visualisations solutions	pp. 248–50
	• assess the effectiveness of the project plan in managing the project	pp. 254–5

9780170440875

OUTCOME	KEY KNOWLEDGE	LOCATION
Unit 4 **Area of Study 2** **Outcome 2**	Cybersecurity: Data and information security On completion of this unit the student should be able to respond to a teacher-provided case study to investigate the current data and information security strategies of an organisation, examine the threats to the security of data and information, and recommend strategies to improve current practices.	
Digital systems	• characteristics of wired, wireless and mobile networks	pp. 263–71
	• types and causes of accidental, deliberate and events-based threats to the integrity and security of data and information used by organisations	pp. 272–5
	• physical and software security controls for preventing unauthorised access to data and information and for minimising the loss of data accessed by authorised and unauthorised users	pp. 275–84
	• the role of hardware, software and technical protocols in managing, controlling and securing data in information systems	pp. 284–9
	• the advantages and disadvantages of using network attached storage and cloud computing for storing, communicating and disposing of data and information	pp. 289–90
Data and information	• characteristics of data that has integrity, including accuracy, authenticity, correctness, reasonableness, relevance and timeliness	pp. 20–3
Interactions and impacts	• the importance of data and information to organisations	p. 298
	• the importance of data and information security strategies to organisations	p. 302
	• the impact of diminished data integrity in information systems	pp. 303–4
	• key legislation that affects how organisations control the collection, storage, communication and disposal of their data and information: the *Health Records Act 2001*, the *Privacy Act 1988* and the *Privacy and Data Protection Act 2014*	pp. 304–13
	• ethical issues arising from data and information security practices	pp. 314–17
	• strategies for resolving legal and ethical issues between stakeholders arising from information security practices	pp. 317–18
	• reasons to prepare for disaster and the scope of disaster recovery plans, including backing up, evacuation, restoration and test plans	pp. 318–27
	• possible consequences for organisations that fail or violate security measures	p. 328
	• criteria for evaluating the effectiveness of data and information security strategies	pp. 331–3
Key skills	• analyse and discuss the current data and information security strategies used by an organisation	pp. 298–302
	• propose and apply criteria to evaluate the effectiveness of current data and information security strategies	p. 302
	• identify and evaluate threats to the security of data and information	pp. 272–5, 302
	• identify and discuss possible legal and ethical consequences of ineffective data and information security strategies	p. 328
	• recommend and justify strategies to improve current data and information security practices	p. 328

9780170440875

Problem-solving methodology

When an information problem exists, a structured problem-solving methodology is followed to ensure that the most appropriate solution is found and implemented. For the purpose of this course, the problem-solving methodology has four key stages: analysis, design, development and evaluation. Each of these stages can be further broken down into a common set of activities. Each unit may require you to examine a different set of problem-solving stages. It is critical for you to understand the problem-solving methodology because it underpins the entire VCE Applied Computing course.

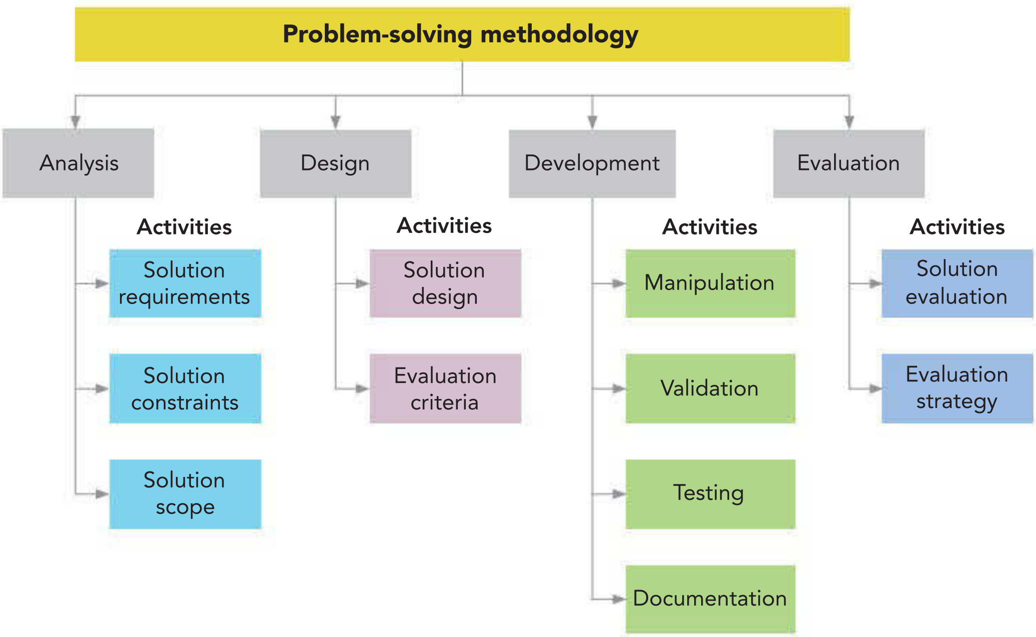


FIGURE 1 The four stages of the problem-solving methodology and their key activities

Analyse the problem

The purpose of analysis is to establish the root cause of the problem, the specific information needs of the organisation involved, limitations on the problem and exactly what a possible solution would be expected to do (the scope). The three key activities are:

1. identifying solution requirements – attributes and functionality that the solution needs to include, information it must produce and data needed to produce this information
2. establishing solution constraints – the limitations on solution development that need to be considered; constraints are classified as economic, technical, social, legal and related to usability
3. defining the scope of the solution – what the solution will and will not be able to do.

9780170440875

Design the solution

During the design stage, several alternative design ideas based on both appearance and function are planned and the most appropriate of these is chosen. Criteria are also created to select the most appropriate ideas and to evaluate the solution's success once it has been implemented. The two key design activities include the following.

1 Creating the solution design – it must clearly show a developer what the solution should look like, the specific data required, and how its data elements should be structured, validated and manipulated. Tools typically used to represent data elements could include data dictionaries, data structure diagrams, input–process–output (IPO) charts, flowcharts, pseudocode and object descriptions. The following tools are also used to show the relationship between various components of the solution: storyboards, site maps, data flow diagrams, structure charts, hierarchy charts and context diagrams. Furthermore, the appearance of the solution, including elements like a user interface, reports, graphic representations or data visualisations, needs to be planned so that overall layout, fonts and their colours can be represented. Layout diagrams and annotated diagrams (or mock-ups) usually fulfil this requirement. A combination of tools from each of these categories will be selected to represent the overall solution design. Regardless of the visual or functional aspects of a solution design at this stage, a design for the tests to ultimately ensure the solution is functioning correctly must also be created.

2 Specifying evaluation criteria – during the evaluation stage, the solution is assessed to establish how well it has met its intended objectives. The criteria for evaluation must be created during the design stage so that all personnel involved in the task are aware of the level of performance that will ultimately determine the success or otherwise of the solution. The criteria are based on the solution requirements identified at the analysis stage and are measured in terms of efficiency and effectiveness.

Develop the solution

The solution is created by the developers during this stage from the designs supplied to them. The 'coding' takes place, but also checking of input data (validation), testing that the solution works and the creation of user documentation. The four activities involved with development include the following.

1 Manipulating or coding the solution – the designs are used to build the electronic solution. The coding will occur here and internal documentation will be included where necessary.

2 Checking the accuracy of input data by way of validation – manual and electronic methods are used; for example, proofreading is a manual validation technique. Electronic validation involves using the solution itself to ensure that data is reasonable by checking for existence, data type and that it fits within the required range. Electronic validation, along with any other formulas, always needs to be tested to ensure that the solution works properly.

9780170440875

3 Ensuring that a solution works through testing – each formula and function, not to mention validation and even the layout of elements on the screen, needs to be tested. Standard testing procedures involve stating what tests will be conducted, identifying test data, stating the expected result, running the tests, reporting the actual result and correcting any errors.
4 Documentation allowing users to interact with (or use) the solution – while it can be printed, in many cases it is now designed to be viewed on screen. User documentation normally outlines procedures for operating the solution, as well as generating output (like reports) and basic troubleshooting.

Evaluate the solution

Sometimes after a solution has been in use by the end-user or client, it needs to be assessed or evaluated to ensure that it has been successful and does actually meet the user's requirements. The two activities involved in evaluating a solution include the following.

1 Evaluating the solution – providing feedback to the user about how well the solution meets their requirements or needs or opportunities in terms of efficiency and effectiveness. This is based on the findings of the data gathered at the beginning of the evaluation stage when compared with the evaluation criteria created during the design stage.
2 Working out an evaluation strategy – creating a timeline for when various elements of the evaluation will occur and how and what data will be collected (because it must relate to the criteria created at the design stage).

9780170440875

Key concepts

Within each VCE Applied Computing subject are four key concepts the purpose of which is to organise course content into themes. These themes are intended to make it easier to teach and make connections between related concepts and to think about information problems. Key knowledge for each Area of Study is categorised into these key concepts, but not all concepts are covered by each Area of Study. The four key concepts are:

1. digital systems
2. data and information
3. approaches to problem solving
4. interactions and impact.

Digital systems focus on how hardware and software operate in a technical sense. This also includes networks, applications, the internet and communication protocols. Information systems have digital systems as one of their parts. The other components of an information system are people, data and processes.

Data and information focuses on the acquisition, structure, representation and interpretation of data and information in order to elicit meaning or make deductions. This process needs to be completed in order to create solutions.

Approaches to problem solving focuses on thinking about problems, needs or opportunities and ways of creating solutions. Computational, design, and systems thinking are the three key problem-solving approaches.

Interactions and impact focuses on relationships that exist between different information systems and how these relationships affect the achievement of organisational goals and objectives. Three types of relationships are considered:

1. people interacting with other people when collaborating or communicating with digital systems
2. how people interact with digital systems
3. how information systems interact with other information systems.

This theme also looks at the impact of these relationships on data and information needs, privacy, and personal safety.

9780170440875

Unit 3

INTRODUCTION

In Unit 3 of Data Analytics, you will use data that you have effectively and efficiently identified and extracted. You will consider the integrity and source of the data and make sure that you have correctly referenced the data using the American Psychological Association (APA) referencing system. You will manipulate this data to create data visualisations and infographics. To do this, you will use databases, spreadsheets and data manipulation software.

You will propose a research question and then collect data to answer this research question. You will use a range of methods to analyse this data. You will use all the stages of the problem-solving methodology (PSM) to prepare a project plan. This will complete the first half of the School-assessed Task (SAT) (Unit 3, Outcome 2). The second half of the SAT will be completed in Unit 4 (Unit 4, Outcome 1).

Area of Study 1 – Data analytics

OUTCOME 1 In this Outcome, you will respond to teacher-provided solution requirements and designs. You will extract data from large data repositories. You will manipulate and cleanse the data and use database, spreadsheet and data manipulation software to present your findings in the form of a data visualisation.

Area of Study 2 – Data analytics: Analysis and design

OUTCOME 2 In this Outcome, you will propose a research question and then collect data to answer this research question. You will use a range of methods to analyse this data. You will use all the stages of the problem-solving methodology to prepare a project plan. This will complete the first half of the School-assessed Task (SAT) (Unit 3, Outcome 2). The second half of the SAT will be completed in Unit 4 (Unit 4, Outcome 1).

iStock.com/gremlin

CHAPTER 1

Data and presentation

KEY KNOWLEDGE

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

Data and information

- techniques for efficient and effective data collection, including methods to collect census, Geographic Information System (GIS) data, sensor, social media and weather
- factors influencing the integrity of data, including accuracy, authenticity, correctness, reasonableness, relevance and timeliness
- sources of, and methods and techniques for, acquiring authentic data stored in large repositories
- methods for referencing primary and secondary sources, including American Psychological Association (APA) referencing system
- characteristics of data types

Approaches to problem solving

- methods for documenting a problem, need or opportunity
- methods for determining solution requirements, constraints and scope
- design tools for representing databases, spreadsheets and data visualisations, including data dictionaries, tables, charts, input forms, queries and reports
- design principles that influence the functionality and appearance of databases, spreadsheets and data visualisations
- formats and conventions applied to data visualisations to improve their effectiveness for intended users, including clarity of message

Interactions and impact

- reasons why organisations acquire data.

FOR THE STUDENT

If you can imagine the sheer amount of data that is generated every day, you might also be able to imagine that there is someone, somewhere, who is looking through a mountain of data searching 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. This chapter will cover the definitions of data and information, the various ways in which data can be acquired and referenced and how to check that data is reliable enough to be used to generate useful information. This chapter will then look at the many types of data visualisations and the design tools that could be used to help plan their use.

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. It covers data types, data integrity and citing references before covering a range of data visualisation tools and their purposes.

The key knowledge and skills are based on Unit 3, Area of Study 1. 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. This chapter, combined with Chapter 2, will form a foundation for the Unit 3, Outcome 1 School-assessed Coursework (SAC). Much of this will be applied when students work on their School-assessed Task (SAT).

9780170440875

What is data?

Data is made up of facts and statistics. Raw facts have no context to them, so you cannot make much sense of them, or give them any meaning. To understand and make meaning of data, you need to process or manipulate it, converting it into something useful: **information**.

Data consists of raw, unorganised facts, figures and symbols fed into a computer during the input process. Data can also refer to ideas or concepts before they have been refined. In addition to text and numbers, data also includes sounds and images (that are both still and moving). Organisations are collecting this data in vast quantities every day. It allows them to plan day-to-day operations, make better business decisions as well as to better understand their customers.

There are several ways in which this data is gathered by organisations. Popular methods include analysing comments on social media, tracking activity on product websites, placing cookies on customer computers and tracking IP addresses.

Organisations can then use data for specific purposes such as targeting customers with advertising tailored to their interests, developing new products and improving existing ones, and even protecting data (for instance, when banks analyse your credit card usage patterns to identify potential fraudulent transactions). There is no doubt that understanding the customer remains a key need of any organisation.

The potential value of data cannot be fully unlocked without processing it into information. Marked ballot papers after an election hold a great deal of data, and thus considerable potential value, but in their raw form, they hold little value. They need to be processed through counting and grouping the ballot papers into their electorates before they become useful information – that is, election results.

Datum is the singular form of data, which is technically plural. Today, nearly everyone uses *data*, the plural form, for both singular and plural.

A cookie is a small file that a web server stores on a user's computer. Cookies typically contain data about the user, such as their email address and browsing preferences. The cookie is sent to the computer when a website is browsed and stored on the computer's hard disk. The next time the website is visited, the browser retrieves the cookie from the hard disk and sends the data in the cookie to the website. Cookies are not viruses because they cannot be executed or run, and they cannot replicate themselves; however, they can be misused as spyware. Cookies can be used to track people, which leads to privacy issues.

FIGURE 1.1 **a** Raw data in the form of ballot papers are **b** grouped and counted to produce **c** election results.

In this chapter, you will learn more about data, data collection and how information is formed and visualised through manipulation. As part of Unit 3, Outcome 1, you will need to collect data and manipulate it using both a relational database and spreadsheet software to create meaningful data visualisations. These manipulations will be covered in detail in Chapter 2.

We will also discuss the data you need to collect, in terms of how it should be treated and manipulated, and how limiting factors involving **constraints** and scope can affect the data.

Data should be gathered from reputable sources, so we will cover how you can acquire data through existing sources, and then measure suitability and integrity by questioning how the data was acquired, such as through surveys, interviews or observation. We will also discuss how to reference those sources properly.

The IP addressing standard – four numbers between 0 and 255 separated by full stops – defines a mechanism to provide a unique address for each computer on a network.

SCHOOL-ASSESSED TASK TRACKER

☐ Project plan	☐ Collect complex data sets	☐ Analysis	☐ Folio of alternative designs	☐ Infographic or dynamic data visualisations	☐ Evaluation and assessment	☐ Finalise report or visual plan

9780170440875

Once you have gathered all of this data, you need to store it, protect it and understand what type of data it is. We will discuss data integrity and how to maintain it through measures such as timeliness, accuracy, authenticity and relevance. This is important because you need to maintain the integrity of the data collected for your Outcome so that your final product can be considered reliable.

Why organisations acquire data

Organisations depend on data in order to function. They use it to keep track of stock levels, client details, employee details, rosters, finances and records of their work. They also use data analytics to make predictions about all aspects of their business.

If an organisation were to lose all of its data, it would suffer greatly. The organisation would not be able to keep track of any finances, it would lose track of its client-base and, if the general public became aware of the data loss, it would suffer a loss of reputation. This would potentially result in fewer clients and the possibility of legal action and closure of the business.

Information needs

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 worried about declining sales), an identified need (park rangers needing 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 create a solution that will meet the needs of the clients or users is the **problem-solving methodology (PSM)**.

Data acquisition

Trusting that data has been collected and stored in a secure manner is discussed on page 20 under the 'Data integrity' heading.

Acquisition is when raw data is gathered from the world outside the **information system**. First-hand, or **primary data**, may be acquired manually, via surveys, interviews or observation or it may be acquired electronically. Electronic acquisition can be completed in many ways: through cameras, people inputting data manually, sensors detecting something such as movement, or it may be acquired through other electronic means (for example, if using a keylogger or scanner).

Data can also be acquired by locating repositories of data that already exist, often online, that has been compiled by someone else or another organisation. If it has not been collected by the organisation directly, or if it has been manipulated or summarised in any way, then it is considered **secondary data**. This form of data can save a lot of time, but you must ensure that the data has come from a trustworthy source. It is important to know how this data was collected, by whom and if there are any reasons to doubt its reliability.

Primary and secondary data

Data that has not been filtered by interpretation or evaluation is called primary data. Often, these are facts that you, the researcher, have collected directly to answer a specific question, but it may also be old data that has never been given proper scrutiny before. The lack of previous interpretation is what categorises something as primary data.

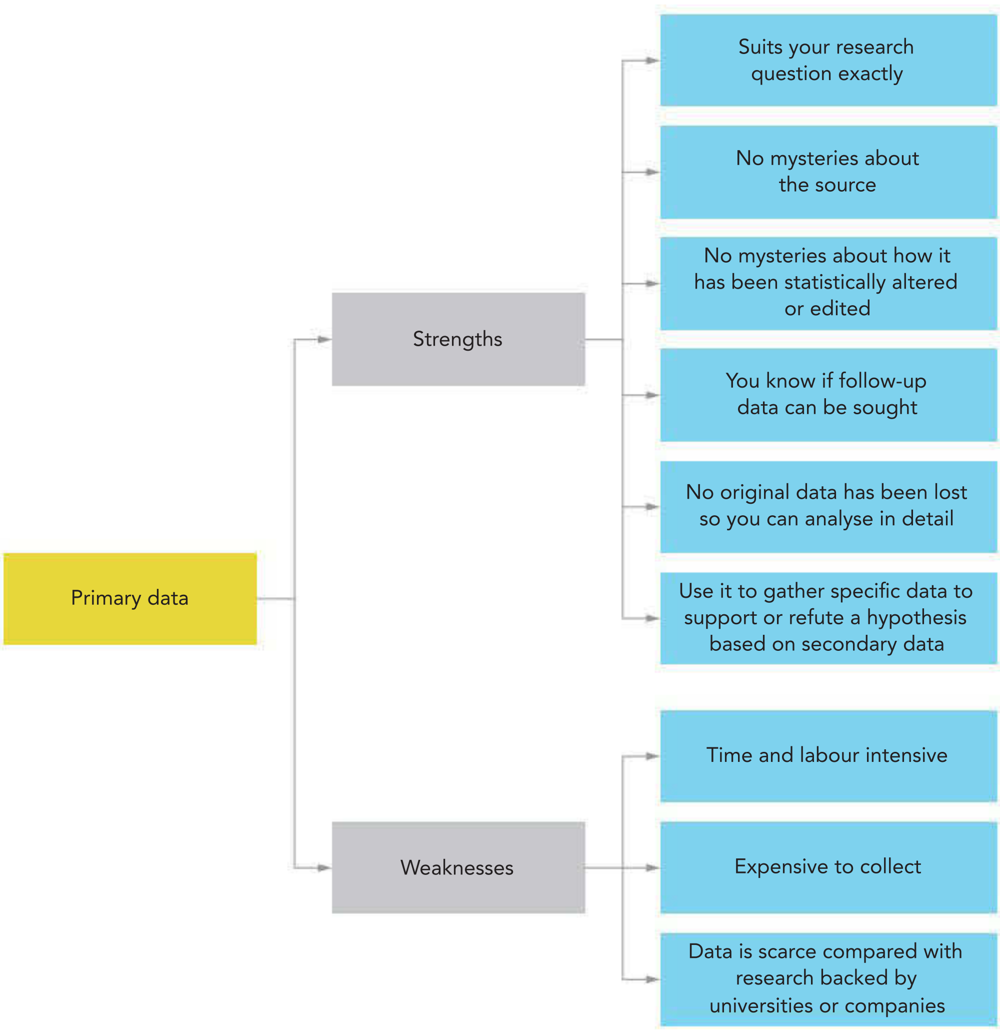


FIGURE 1.2 Strengths and weaknesses of primary data

Journal articles, results of questionnaires, diaries and letters, emails, internet postings, speeches, audio and video recordings (if unedited), official documents and photos may all be primary data. Therefore, when *you*, as the researcher, need to interpret or evaluate unedited data that you are using in your Outcome, it is primary data.

Researchers also collect (new) primary data using interviews, focus groups, surveys, experiments, observations and measurements.

Secondary data differs from primary data because it has been collected and interpreted by someone *other* than the researcher. The collectors of the data could be other researchers, or government departments, or any of a variety of sources: encyclopaedias and other books, biographers, conductors of polls and surveys, journalists, newspapers and magazines, the Australian Bureau of Statistics, internet posts, databases and so on. When using secondary data, it is especially important to consider whether both the data and its interpretation are from a reputable source. Therefore, when you collect data for your Outcome that has been interpreted by someone else already, you should consider this to be secondary data.

9780170440875

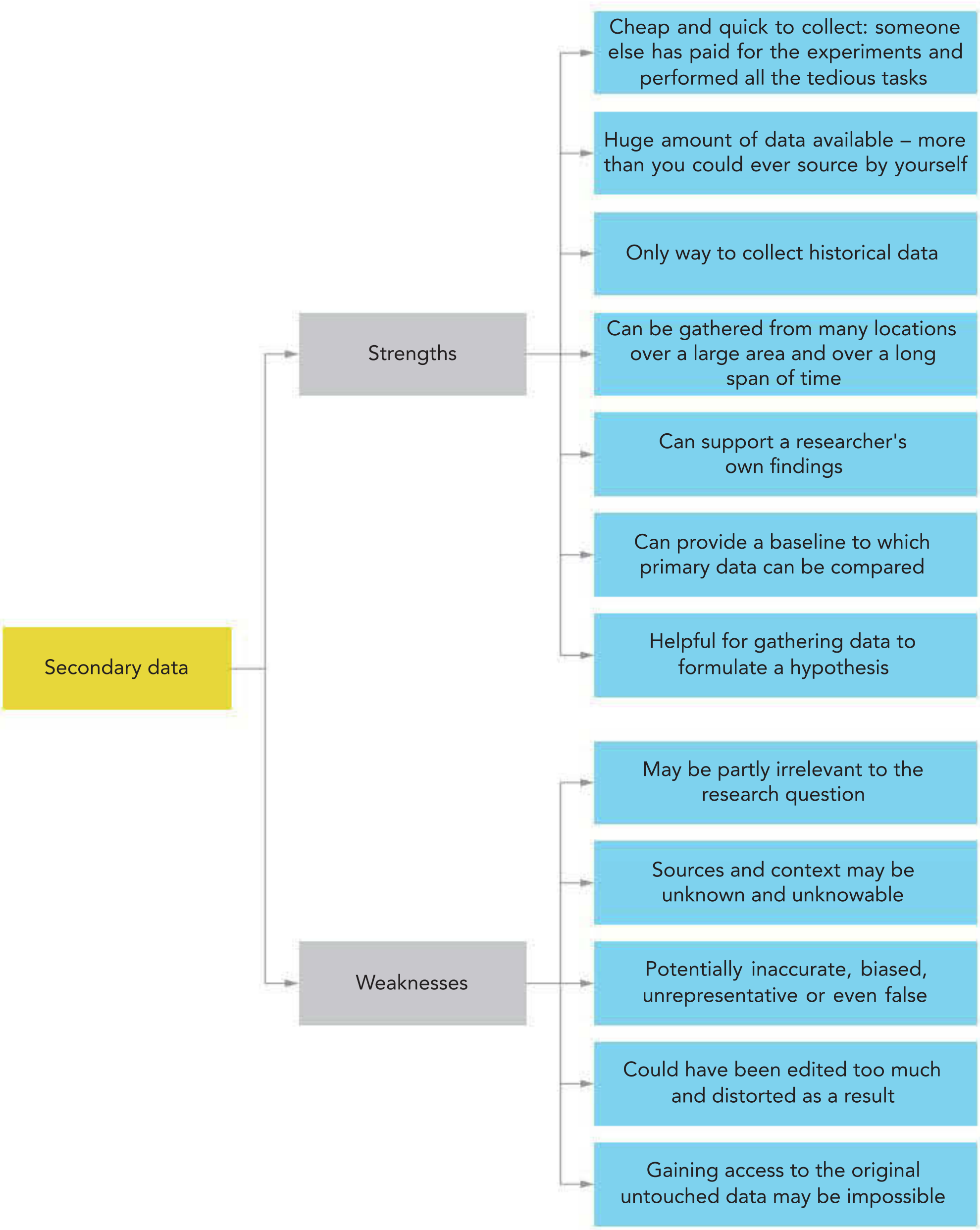


FIGURE 1.3 Strengths and weaknesses of secondary data

Techniques to collect data

Ideally, data should be collected both **efficiently** (not wasting time, cost or effort) and **effectively** (in a way that ensures data is complete, usable, accurate and current). How the data is collected can affect the quality or integrity of the data. Data can be collected by using surveys, interviewing participants or observation. Each method has its own benefits and drawbacks; you need to know how the data you are using has been collected in order to understand the impact it will have upon the quality of your first SAC.

Techniques to collect data will also be covered later when you begin working on your SAT for Unit 3, Outcome 2.

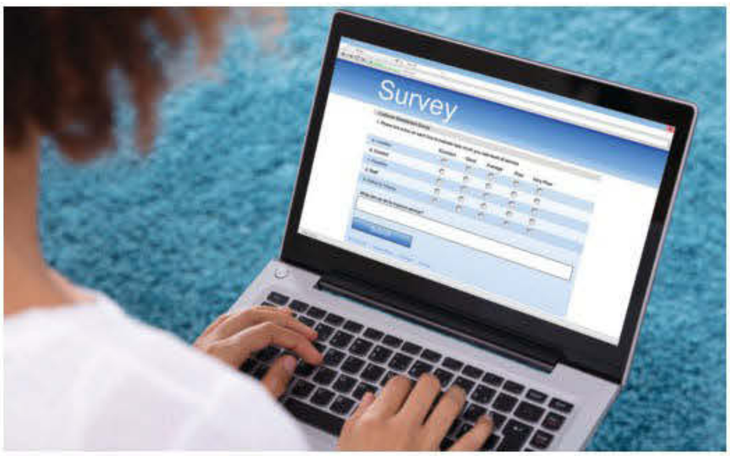
Survey

Surveys are a fast and relatively cheap way to gather large amounts of data and feedback. They can be administered in many different ways – online (having users enter their own data makes things much easier and quicker), on paper (circling numbers, ticking boxes, writing short responses) or verbally in person or over the telephone. The questions in a survey remain identical for each person completing it, so that if any further clarification is required, this cannot be done, especially if the survey is anonymous. If it becomes apparent that extra questions need to be added, it is probably too late to do so – once produced, a survey is fixed.

For example, if a respondent misunderstood a question, or if the response given contradicts an earlier response, then the quality of data might be affected. There is no way of gauging whether or not people are being honest in their responses as with a face-to-face interview, which might allow for tone and body language to make points of view clearer.

Question types are limited to:

- checkbox for Boolean data: yes/no or true/false questions only
- scaled responses: Likert scales ask respondents to select how they feel about a particular statement, asking for a number from 1–5 (or any scale)
- closed questions: asking for a response from a set number of options
- open questions: respondents give worded responses with no limitations; this can give more detail, but it is more difficult to use this data when collecting from a large number of people.



Shutterstock.com/Andrey_Popov

FIGURE 1.4 Conducting surveys online saves time and effort because they avoid added manual data entry.

Interview

Interviews take place with two or more people in real time. They can be conducted face-to-face, via video or telephone, and with individuals or small groups. Like surveys, the questions are usually written in advance so that responses can be compared to get the big picture from the data collected. Unlike surveys, respondents can request clarification if they do not understand a question, or the interviewer can probe for more detail if they think that it is appropriate. For example, an interviewer might ask the respondent to provide an example or ask why they said something. Interviews are more costly, in that they cannot be deployed to as many people or as quickly as surveys, but the quality of data is usually better.

9780170440875

FIGURE 1.5 Interviewing allows for follow-up and clarification of questions.

Observation

Observations are the most costly data-collection technique when collecting data about real-time events or existing processes. They involve watching and taking notes in real time as an impartial observer. This is far slower than the previous two techniques, but results in far richer and more genuine data. This kind of data collection is better suited to studies that do not require large amounts of data. Observations are generally a good idea to combine with surveys and interviews to allow more information to be collected. For example, if a respondent's body language in an interview indicated that they were unsure of their answer, the interviewer could request more detail around the response to gather more usable data.

Sensor

Sensors or data loggers are devices used to detect characteristics of the environment around them. This may include temperature, humidity, light levels, motion, touch, and the amount of gases in the air. Sensors are always connected to other electronics that can interpret the electrical signals they generate. Sensors can be connected together to provide an overall snapshot of a particular environment (as with a weather station) and even set-up to transmit

FIGURE 1.6 Sensors can be used to measure air quality.

data from remote locations (including other planets). It is important that data collected from sensors is validated to reduce anomalies and stored securely to reduce the possibility of data loss.

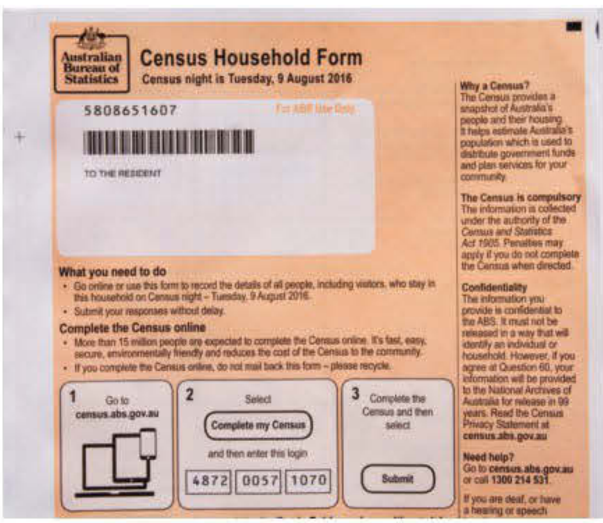
Sensors collect data electronically. Once installed, they can monitor and transfer data to storage without human intervention. They can operate 24 hours a day, 7 days a week, and they can gather data on weather, movement, traffic speeds on roads, levels of light, noise or pollution, or numbers of people or cars entering or leaving facilities, or almost anything else.

Methods applied to specific data collections

The following data repositories have collected vast amounts of data in various ways. The organisations and their data will be discussed later in this chapter. Considering the sheer volume of data being collected by organisations, it would be impossible for a human to record it either efficiently or effectively. Therefore, there are several automated processes commonly used to gather data. Common types of data to be gathered include census, **Geographic Information System (GIS)** data, sensor, social media and weather data.

Census

Every five years, the Australian Bureau of Statistics (ABS) conducts the Census of Population and Housing. This is one of the largest data collecting activities in Australia. It is designed to provide a demographic snapshot of Australian society. In previous years, households were asked to complete a paper survey booklet containing questions about various characteristics of the people living in a house on a particular night (Census Night). The 2016 Census was the first time census data could be entered via an online portal. Households were provided with login details for authentication. Responses to questions were validated electronically and deidentified. While the collection of census data was easier in 2016 when compared with the paper-based forms, the data entry process was more vulnerable to interference, as was demonstrated when hackers caused the portal to be shut down temporarily. Those who were unable to complete the census online were given extra days to complete the task. It is a legal requirement for each household to complete the Census, and the honest, accurate completion of nearly every question is required.



Alamy Stock Photo/chris24

FIGURE 1.7 The ABS census provides a snapshot of Australians and Australian households every five years.

9780170440875

Weather

The Bureau of Meteorology (BOM) uses sensors to collect data on all aspects of weather and climate. Data is automatically gathered and stored and is made available on the BOM website. The sensors are not subject to personal bias or human error, and unless they malfunction their accuracy and reliability is very high.

The National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and Space Administration (NASA) in the United States also use sensors to gather climate data, which is available from their websites.

Social media

Many social media sites generate data from their users. This data, sometimes only available through subscription, summarises how far the account has reached, and how many views and interactions have happened. It is not always easy to locate the raw data.

Social media sites also take part in data mining – storing data about its users. This data is gathered in order to learn about users for, among other things, targeted advertising.

Interaction with popular social media platforms such as Facebook, Twitter and LinkedIn provides a range of measurable data elements. This data comes from interactions such as clicks, comments, likes, shares and conversations. Social media users can also be broken down further by location and language preferences. This data can reveal the success of a marketing campaign or evaluate customer impressions of a product.

Geographic Information System (GIS)

Geographic Information Systems use sensors to record data about Earth. They capture the data, store it, carry out manipulation to create useful information and then present it in easily understood ways. A GIS stores data on geographical features as well as their characteristics. Features are usually classified as points, lines or areas. Data might also be stored as images. City data on a map might be stored as points, road data as lines, and boundaries as areas, while aerial photographs or scanned maps could be stored as raster images. The ABS presents a lot of their data as GIS data.

Data sources

Finding a relevant **data source** takes thought, judgement and care. A great deal of data is available online. The trick is to find the data that is worth using and to use it correctly. While Chapter 3 deals with the forming of a research question, this chapter is about the nature of data itself.

It is important that you investigate both primary and secondary data sources. This will be discussed in much more detail as you prepare for the next Outcome.

The source of data is very important. Treat data without an identified source with care and take necessary steps to resolve matters of **authenticity**. When the source of the data is known, it is more reliable. Without this, your trust would be blind because you would not be able to contact the data's creator – the data could even be a work of fiction. Putting complete trust in anonymous data is especially unwise. For example, if you find survey results on social media, it is important to check to see if it is supported by substantiated facts and that the data has not been made up by somebody.

However, if you know how data has been collected, statistically manipulated, edited and/or abridged, you can interpret the information more wisely. A reputable data source with authority is more likely to provide high-quality data. Be wary of data given by people who *are* experts, but *not* necessarily experts in the relevant field. The opinion of a champion footballer on politics, or an actor on climate change, are no better than any other person's opinion.

There are hundreds of data sites in Victoria and Australia on many topics. Choose a question that interests you. Search online for key words relevant to your question and add 'data set'.

You will probably find data that will inspire many reasonable and interesting research questions.

Data from government organisations

Data that is collected by organisations is sometimes made available to the public, often online. Many government bodies, including Australian federal and state governments, make some data sets available. For Unit 3, Outcome 1, you will be using some of this data. It is important to know about these organisations since you need to be able to trust them as data sources.

Australian Bureau of Statistics (ABS)

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 to agriculture, sporting facilities and crime statistics. In fact, the ABS website provides an endless supply of open data, downloadable and ready for manipulation.

Australian Bureau of Statistics

DataVic

DataVic offers public data sets from many locations, groups and topics, including a national public toilet map, overseas arrivals and departures, taxation statistics and even Victorian ice-skating centres.

DataVic

Commonwealth Scientific and Industrial Research Organisation (CSIRO)

The CSIRO is an Australian scientific organisation that collects and stores scientific data. It manages national scientific research and focuses on challenges that face both Australia and the world. The CSIRO collaborates with many leading organisations worldwide to solve challenges using innovation around science and technology, and is committed to shaping the future to improve communities, economies and the planet.

CSIRO

Bureau of Meteorology (BOM)

The Bureau of Meteorology is Australia's national weather, climate and water agency. The BOM collects a wide range of climatic data in order to make regular forecasts, issue warnings and offer advice with regards to the weather. Much of the data that it collects is available for public use.

Weather data is collected by a range of devices including thermometers (temperature), radar systems (used to create maps of rain and snow and measure rain cloud movement), barometers (air pressure), rain gauges (how much rain falls), wind vanes (wind speed and direction), transmissometers (visual range), and hygrometers (humidity and water vapour). These devices work in real time to provide current weather information as well as assist with weather prediction.

Bureau of Meteorology

9780170440875

The Bureau of Meteorology operates a large number of weather stations throughout Australia that have a combination of these devices. Australian weather data is also shared with other countries, and vice versa, in order to establish global weather patterns. Analysis of weather data is normally done on a supercomputer.

NASA

FIGURE 1.8 The BOM's climate data online

National Aeronautics and Space Administration (NASA)

The United States' National Aeronautics and Space Administration (NASA) website has large repositories of data available to download. Its data and research are focused not only on space, but geographic and climate data for Earth as well. For example, much of the data presented about global warming and pattern trends worldwide comes from NASA.

THINK ABOUT DATA ANALYTICS 1.1

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

Not-for-profit organisations

Organisation for Economic Co-operation and Development (OECD)

The Organisation for Economic Co-operation and Development (OECD) is a group of countries working together. The aim is to use its wealth of information on a broad range of topics to help governments foster prosperity and fight poverty through economic growth and financial stability. The OECD helps to ensure the environmental implications of economic and social development are taken into account. The OECD collects global data, by country, across many areas deemed important to making lives better for people, including gross domestic product (GDP), education levels, mortality rates, health data, financial data and much more. Investigate Gapminder online (see weblink) to engage with some of their data.

Gapminder
Gapminder is a not-for-profit organisation that specialises in data visualisation of large data sets from reputable sources including OECD, WHO (World Health Organization), UNAIDS, International Labour Organization, UNESCO and the United Nations. You can choose your own data sets and create stunning visualisations.

Other collections of data

Social media

Social media sites have in-built tools to help users locate data to track their impact. 'Likes' and 'shares' are attached to each post, as are comments. Figure 1.9 shows some Twitter feedback on an individual tweet.

Sensor data

Sensor data is anything that is collected via sensors. These can include road (traffic) cameras, video surveillance, toll data, Bluetooth capture boxes or something that you set-up yourself that you could use to potentially aid your research in your Unit 3, Outcome 2 SAT.

This type of data is collected and often shared by a wide variety of organisations, from government sites such as those listed previously to commercial organisations and private citizens. It is up to you as a researcher to confirm the reliability of the data you are using.

Impressions	765
Total engagements	93
Detail expands	38
Media engagements	19
Likes	19
Profile clicks	10
Replies	2
Retweets	2
Hashtag clicks	2
Link clicks	1

FIGURE 1.9 An example of statistics available from Twitter to measure tweet impact and reach

Geographic Information System (GIS)

Referred to as GIS, the purpose of the Geographic Information System is to handle spatial or geographic data. Typically, GIS can overlay multiple sets of seemingly unrelated geocoded data (data that includes elevation/altitude, longitude and latitude) according to time of occurrence to show events in location. These are used to create maps that can show distribution of specific items or events. For example, by recording the home address of patients suffering from particular diseases, doctors can establish if there is a link with a particular location by plotting these on a GIS map.

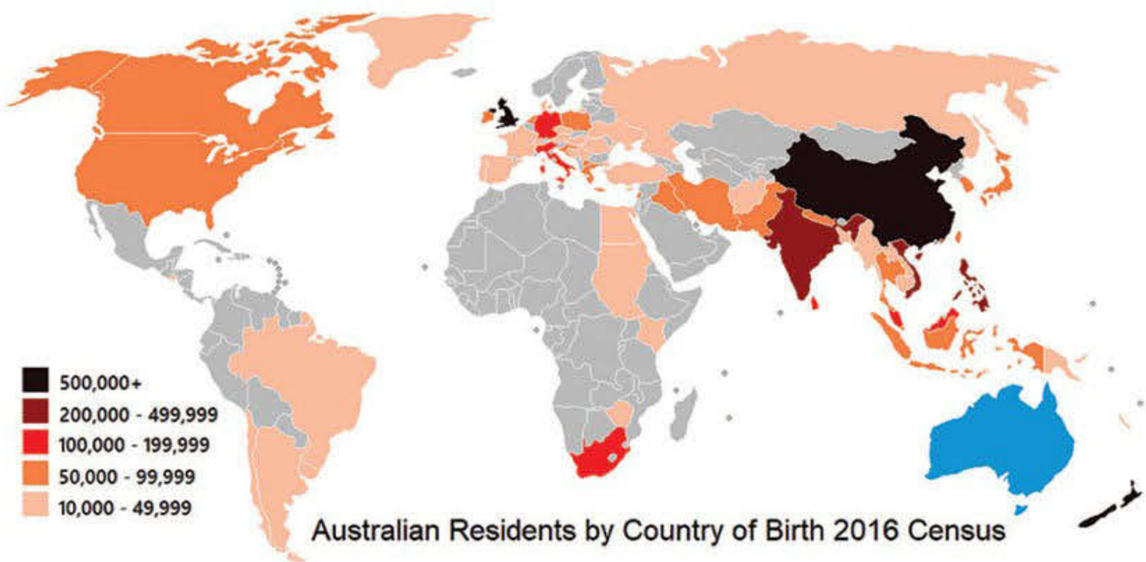


FIGURE 1.10 This example shows how GIS visualisations can combine map data with census data from the Australian Bureau of Statistics.

9780170440875

Referencing data sources

You must acknowledge all sources of data and information you use in your work. This means not only direct quotes, but also ideas, summarising and paraphrasing. There are numerous referencing styles available. All use a brief identifying **citation** in the body text (either in brackets or as a footnote) and a matching detailed reference in a reference list and/or a bibliography. A reference list only includes the works cited in the body text. A bibliography also includes other texts that were consulted and may be of interest to the reader but were not referred to in text. A bibliography can also be annotated. The sources of **intellectual property (IP)** that influence your work in Data Analytics should be acknowledged using the **American Psychological Association (APA)** referencing system. Note that APA style uses a reference list and not a bibliography.

This referencing style is not just applicable to this Outcome; it is also used for the first part of the SAT (Unit 3, Outcome 2). Referencing in this way ensures that:

- it is clear that you are not claiming another author's work as your own
- readers can find the original source to get more information
- **copyright** 'fair dealing' legalities are observed
- moral rights are observed.

It is wise to start recording your sources as you find them, rather than at the end of the Outcome. Having to locate sources again to cite them may be difficult and time-consuming.

If you fail to acknowledge someone else's intellectual property in your work, you may be accused of, and be guilty of, plagiarism. **Plagiarism**, which is when you pass someone else's original work off as your own, is very serious. It can have repercussions beyond embarrassment: you may find yourself in academic and legal difficulty. Your school or university may take disciplinary action. You may even face prosecution or civil legal action under copyright laws.

APA style guide

The American Psychological Association (APA) referencing system is widely used and is one of the most common styles for referencing. Citations appear within the text showing the author's last name, the year of the publication and the page number of a quote. These are usually separated by commas and are placed within parentheses following the text. For example:

> Despite the fact that fruit and vegetables seem, at first glance, similar enough in terms of dietary requirements, and tend to be used interchangeably, the topic of diet can be contentious. Fruit tends to have higher levels of fructose than vegetables and in general terms a fruit can be a vegetable, but a vegetable cannot be a fruit. (Redman, 2018, p. 231–6)

The corresponding source details appear in a reference list at the end of the work. Microsoft Word can easily be used to keep track of your sources. You just need to fill in the boxes and it can automatically create a reference list or bibliography for you. Tables 1.1 and 1.2 demonstrate how to set out simple citations in your reference list.

9780170440875