Student book answers

Chapter 7 Data analysis and visualisation

Test your knowledge

INFORMATION NEEDS AND DATA VISUALISATIONS

1 If individuals or an organisation requires information that is not already being provided, then an information need exists.

2 An information need may arise due to:

- an existing problem
- identified need
- an opportunity.

3 A data visualisation may be required if more meaningful information is required from large data sets.

SOURCES OF AUTHENTIC DATA

4 Primary data sources involve collecting the data directly or firsthand from stakeholders. Methods used to collect primary data include interviews, observations and surveys. 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.

5 Interviews, surveys and observations

6 The advantage of a survey over an interview is that they can be quickly developed and then sent to a large number of people. This allows feedback from a greater sample size, increasing the likelihood of the result being free from bias.

7 Newspapers, television and websites

8 Responses may vary, but a prominent issue relating to secondary data sources is its reliability. Is the data accurate, relevant and timely? Some secondary data sources may contain data that is not accurate or is misleading.

9 Traditionally, government departments have collected vast amounts of data on a range of topics to meet their needs over a long period of time. Therefore, this makes them a major source of data sets in Australia.

10 The Bureau of Meteorology is Australia's national weather, climate and water agency. It provides weather forecasts, warnings, monitoring and other weather advice.

11 The Victorian government data directory contains data sets collected by Victorian government departments, whereas Data.gov.au contain data sets collected by federal government departments.

12 The ABS is a specialist department. Its purpose is to collect data and produce statistics on a range of topics. This data is available from the ABS website and also the Data.gov.au website. The Data.org.au website also contains data sets collected by other federal government departments.

INTEGRITY OF DATA

13 Data integrity refers to the quality of the data.

14 Accuracy, timeliness, authenticity and relevance

15 Electronic validation techniques can be used when data is collected to improve the chance of the data entered being accurate. Validation techniques include dropdown lists, radio buttons and checkboxes.
16 Timeliness refers to the collecting data in a reasonable timeframe, so that it will be useful enough to produce relevant information. Timeliness can also refer to the information produced, where that information needs to be available in time to be used in decision making.
17 Genuine, original and trustworthy
18 Relevant data would include data that is connected to the purpose of the visualisation. If a visualisation was being developed related to the height of students, then a relevant data set would contain a list of student heights.

DATA TYPES AND STRUCTURES

19 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.
20 If using an example from this chapter, any four of the following: character, text (string), integer, floating point (decimal) or Boolean. Accept examples from other chapters.
21 Both can hold any value (number, letter or symbol), but a character can only hold one single value (for example, %) while a string can hold two or more symbols.
22 Both can hold numbers, but an integer can only hold a whole number while a floating point can store decimals.
23 Boolean would be most appropriate. Boolean is a logical data type that can only store one of two values (true or false) so is suitable for yes or no questions. In addition, Boolean uses a small amount of storage space, so it is an efficient manner of storing data.
24 An array is a data structure that (generally) holds a number of different data items that are of the same data type, such as all integers. A record data structure holds a number of data items that consist of different data types, such as an integer, string, floating point and Boolean.
25 Files are effective because they can hold large amounts of data. They can also store data independently of the software required to view the data. For example, a Word file can store an essay even though Word has been closed.

TYPES AND PURPOSES OF DATA VISUALISATION

26 A data visualisation is the presentation of data in a graphical format to help identify patterns and relationships between the data items.
27 Line, column and pie
28 To provide information in the form of a visualisation of a geographical area or location
29 They allow an area or location to be brought to life by overlying a data set with mapping information.
30 Network visualisations are used to represent relationships between different data items or data sets.
31 Timeline data refers to the tracking of events in the time they occurred. Time series data refers to the tracking of an item over time.
32 Examples include organisational charts, structure charts, tree charts and mind maps. Other examples may be acceptable.
33 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).
34 Matrix visualisations often divide 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.

ANALYSIS STAGE

35 Development of solution requirements, solution constraints and setting out the scope of solution
36 Solution requirements are what the user wants from the solution.
37 A functional requirement is a specific set of tasks that the software must perform. For example, the spreadsheet will calculate the average price.

Non-functional requirements are other elements the user wants the solution to have that is not directly related to what the software can do. For example, the solution needs to have a simple user interface.

38 A functional constraint is a constraint that applies that applies to a functional requirement. A non-functional constraint is one that applies to a non-functional requirement.
39 Constraints can restrict what solution requirements can be included in a solution. For example, a lack of software available to develop a data visualisation, may restrict.
40 The scope outlines what will and will not be included in the solution. This helps stop miscommunication and disagreements between stakeholders occurring after the solution is finished.

DESIGN TOOLS

41 While the analysis stage is all about what the solution needs, the design stage is all about how the solution will function.
42 Designing the solution using design tools and creating evaluation criteria
43 The purpose of a design tool is to represent how the solution will either appear or function.
44 Appearance design tools show a representation of how the solution will look when completed.

Functionality design tools show a representation of how the solution will work or operate when completed.

45 Layout diagram and storyboard
46 By creating alternative design ideas, the merits of each idea can be discussed, with the best idea (or combination of ideas) being selected to be developed. This will assist to create the best solution to solve the information need.
47 An IPO chart is a functional design tool used to identify the data (input) required to create the information (output) needed.
48 A flowchart can be used to represent the series or steps a user needs to complete a data visualisation. The answer given for this question should ideally be relevant to this chapter, but accept answers relevant to earlier chapters.

FORMATS AND CONVENTIONS

49 A format is a way that data and information can be presented. Examples of a format include a document, graph, table, video or visualisation.
50 A convention is a set of general rules that are followed when using particular formats. Examples of a format for a table include subheadings are bold and centred, text is left aligned and numbers are right aligned.
51 Responses will vary. The following four options are an example of four conventions.

- Display the title of the visualisation clearly
- Use key or legend to identify different data items
- Include the name of the author and sources of data
- Identify the units of measurement used

SOFTWARE TOOLS AND FUNCTIONS

52 Examples include spreadsheet software, Google Charts, Tableau Public, OpenHeatMap and Watson Analytics
53 A motion chart can be used to develop time visualisations.
54 Both tools focus on creating map-based visualisations.

FILE FORMATS

55 You can save data created in software applications as files. You can store the data files independent (outside) the software. Later, the data can be read in from the file back into the application.
56 XLM/XLMS files, WMS files and GIS files
57 WMP files, which contain mapping data, can be used in conjunction with GIS files, which contain data based on geographical locations. By using data from both files together geospatial visualisations can be created.
58 A comma-separated value (CSV) file is a plaintext file that stores different data items separated by commas.
59 An API tool allows live data to be collected from an external website.

EVALUATING VISUALISATIONS

60 Evaluation is checking that the working, complete solution meets the information need of the user.
61 Evaluation should take place after a solution's development and implementation. A period of time needs to pass so that the users of the solution have become familiar with the solution before assessing it.
62 Evaluation strategy, evaluation report
63 The evaluation criteria created in the design stage are then used again in the evaluation stage.
64 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.
65 Surveys, observation, interviews, reviewing solution output, checking download speeds, counting website hits, reviewing error logs or timing how long it takes users to complete tasks are all acceptable answers. Accept any three.
66 The purpose of an evaluation report is to state whether each evaluation criterion was achieved or not. The judgement would be based after the data was collected and the techniques applied (that were identified in the evaluation strategy). Finally, the criterion would be judged successful or not depending on the justification that was also developed in the evaluation strategy.
67 If all evaluation criteria are achieved, the solution is deemed a success.

Apply your knowledge

Responses and visualisations will vary depending upon each individual student's selections.