Student book answers

Chapter 1 Data analysis

Test your knowledge

Data and information

1 The information produced will be incorrect and therefore unreliable.
2 Data must be accurate to be usable. If data is not accurate, it will not produce correct information.
3 Usable data is relevant, accurate, free from bias and reliable.
4 Transcription error
5 You should check primary sources against the original data.
6 You should process data while it is current because you should not make decisions based on outdated data. For example, a decision to take out a mortgage should not be based on interest rates that are more than a week old, and stock prices need to be continually updated so that potentially huge financial decisions are not based on out-of-date data.
7 Vested interest, timing, small sample size, bias through sorting and bias through graphic information
8 Qualitative data is hard to measure because it is based on subjective data collection methods such as interviews, focus groups, interviews and observation. It is concerned with experiences, beliefs, feelings and opinions.

   Quantitative data is measurable and specific. It verifies theory using statistics and largely numerical data. Examples of quantitative data include questionnaires and experiments.

9 You can gather quantitative data using questionnaires, experiments, focus groups and observation. Questionnaires in particular can capture a large sample size and may be analysed using software such as Excel. Having a large sample size provides statistical validity, and the accurately reflects the population. Data is then interpreted, relationships identified and findings communicated.
10 Responses will vary. See APA for more details: http://www.apastyle.org/learn/index.aspx
11 You should obtain permission when collecting data because of privacy laws, such as the Australian Privacy Principles (APPs), to establish boundaries and to ensure that participants can make an informed decision to participate.
12 Consent forms help to ensure that all participants have informed consent. Informed consent means that, before agreeing to participate in research, participants are fully aware of what the research involves, the time required from them and the possible risks that may arise. The form also ensures that participation is voluntary.

Digital systems

13 Encryption is important because it is one way to ensure that data is secured, so that only authorised people, in possession of a secret ‘key’, have access to it.
14 A username usually is identifiable as belonging to a particular person. Generally, you can easily remember your own username. Usernames are uniquely assigned to users. Passwords are placed on files and/or folders to prevent unauthorised people from accessing the data.
15 To maintain high levels of security, passwords should be at least digits long, include non-alphabetical characters, not easy to guess and changed every month. Responses will vary for examples of a very strong password.

16 Any of the following responses are acceptable.

- A full backup copies all of the files from a device to a storage medium. It can take considerable time and is usually performed once over a time period (such as a week, fortnight or month).
- A differential backup copies only those files that have been changed since the last full backup. Restoration of data would involve restoring files from the full backup and then from the differential backup.
- An incremental backup is similar to a differential backup, the difference being that it uses more than two backup media, while a differential backup uses only two media. An incremental backup only copies files that have been changed since the last incremental backup. It is the most complicated strategy from which to restore files since it requires restoration from a full backup and then from a series of incremental backups.

17 A full back up copies all files from a device to a storage medium. An incremental backup only copies files that have been changed since the last incremental backup.

18 You should secure data during research to ensure that the integrity and identification of the data is not compromised.

Interactions and impact

19 The APPs protect the individual's data and how it is handled; they were devised to set out the standards, rights and obligations for collecting, handling, holding, accessing, using, disclosing and correcting personal information.

20 The *Privacy Act 1988*, amended by the *Privacy Amendment (Enhancing Privacy Protection) Bill* in 2012, which came into effect in 2014.

21 To maintain privacy; that is, so a single person or entity cannot be identified

22 An ethical dilemma occurs when there is a conflict between two equally desirable principles. This is especially the case when the consequences of the action are open to debate or interpretation. For example, when doctors are requested to send patient medical details to a central authority for a nationwide health study, parents of young children have a dilemma in that they must weigh up their child's right to privacy against the benefits to society.

Apply your knowledge

Street traffic

1 a Does Arthur's street need traffic management?

b The number of cars that use his street, the time, the day, and direction they are travelling; comments made by residents

c Observation and interviews

d Responses will vary, but students should indicate that observation will be required because there is no other way to monitor the traffic, and interviews will be effective to gather input from residents. Questionnaires could be used for the residents, but it is more likely that Arthur will gather responses from his neighbours if he tries to get interviews instead.

e He could keep notes from his observations in a locked filing cabinet. He could keep the data from the interviews on his computer secure as long as only he has the username and password.

f Arthur does not need permission to observe the cars. He should get consent from the residents that he intends to interview so they are aware of what he is doing, and that they are participating on a voluntary basis.

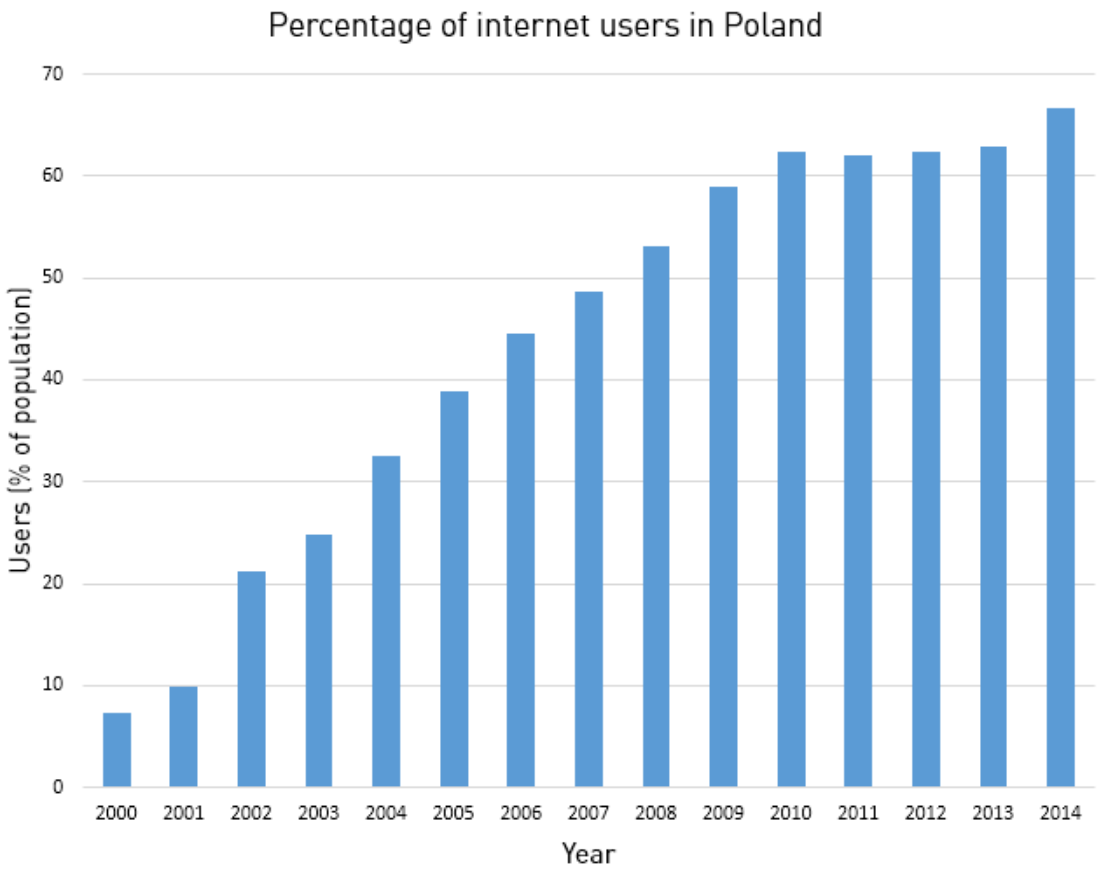
g Arthur could use a spreadsheet to assist with the statistical representation of his observations. In terms of the interview, he could transcribe the interviews and then look for patterns.

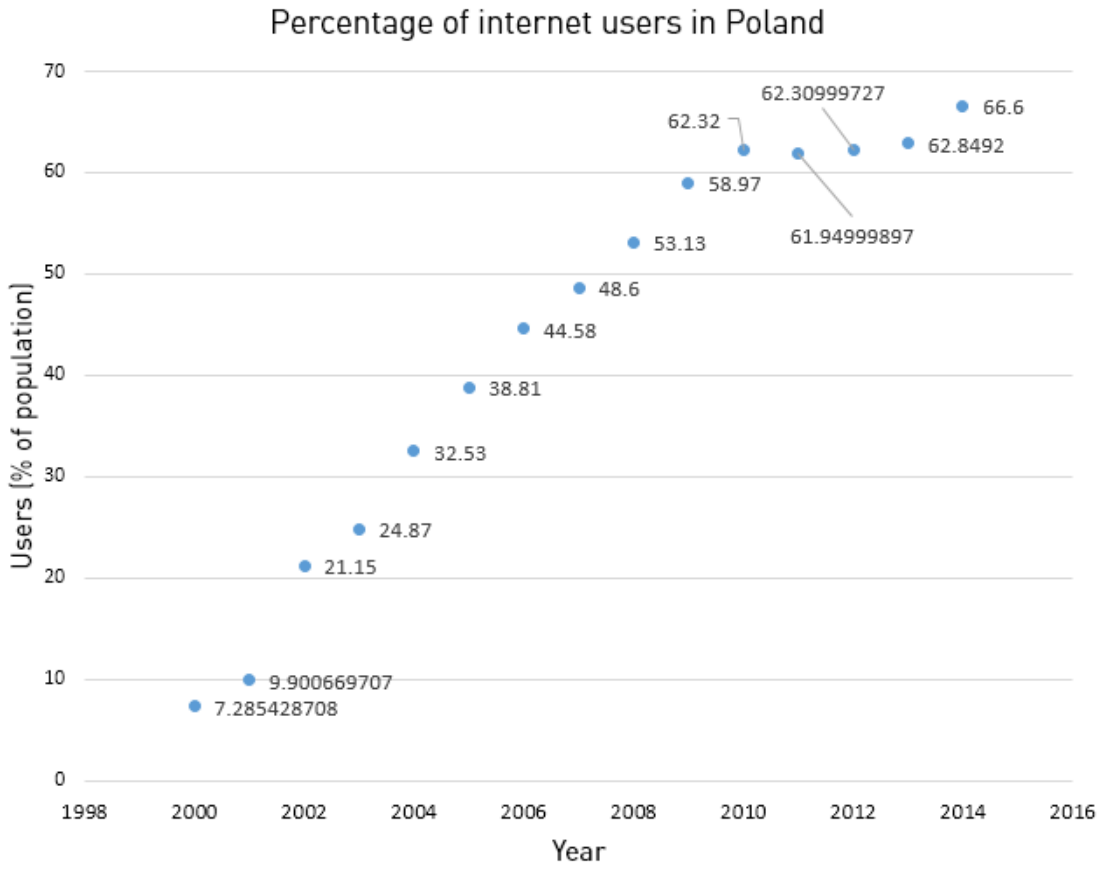
h Arthur is looking for patterns that suggest that there are more cars travelling in one direction in the morning peak, and at what time the cars ease off, what happens in the middle of the day, and what time the evening peak begins and ends and the direction the cars are travelling.

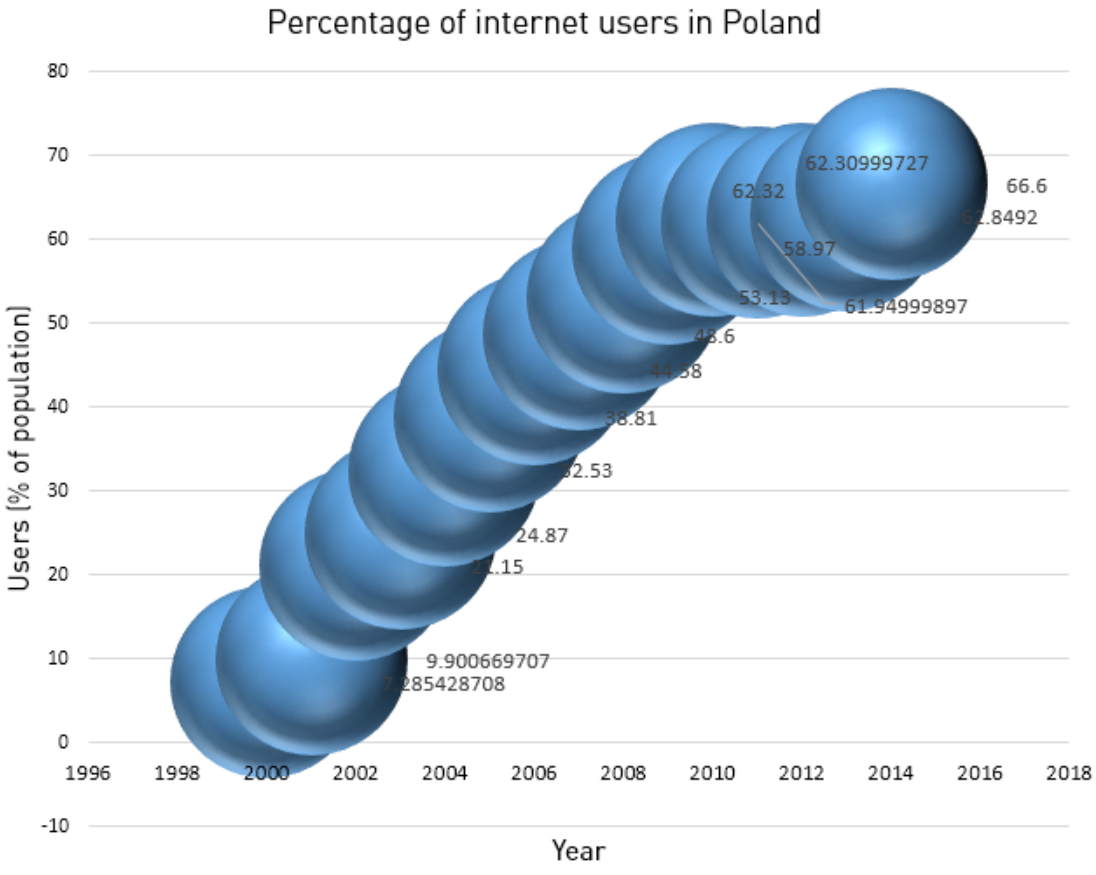
i Arthur can present his findings in a report with graphical representations to make it easier for the local council to interpret his findings.

Internet usage

2 **c** Students should filter the data to choose the data from one country only, and a sensible year span, such as 2000–2014, or they will be left with too large of a data set to create a meaningful or useful set of statistics. The following examples use the country Poland and the years 2000–2014. The scatter diagram and the bubble chart show the values but the column chart does not. Ensure that students are marked on whether they adhere to conventions, such as including a chart title, axis titles and so on.




Percentage of internet users in Poland
Users (% of population)
Year
70
60
50
40
30
20
10
0
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
62.30999727
66.6
62.32
62.8492
58.97
61.94999897
53.13
48.6
44.58
38.81
32.53
24.87
21.15
9.900669707
7.285428708


Percentage of internet users in Poland
Users (% of population)
Year
80
70
60
50
40
30
20
10
0
-10
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
62.30999727
66.6
62.32
58.97
53.13
61.94999897
38.81
24.87
9.900669707

d The scatter diagram best depicts these statistics because it is easy to see the growth in the internet. The bubble chart is not at all suited to this data set. The students should return a bubble chart that looks bad and this is okay. Part of the reason to do a bubble chart is to show it is the wrong type of graphic for this task. The figures are too close together, which makes the bubbles overlap, and there are not enough sets of values for comparison. The overall picture is unclear.