

Solutions

Each answer shows the steps, then the **result**.

2.1

- data that **describes** other data
- for a song: the title, artist, length, date recorded or file size

2.2

any one:

- make the case the **same**
- **trim** extra spaces
- **merge** codes so that all four become one value, e.g. "China"

2.3

- data are **raw** values; information is the **meaning** or knowledge obtained by processing them

2.4

- a **third factor** may affect both, or it may be chance; the data only show they change together

3.1 B

- metadata describes the photo; the pixels are unchanged

3.2 C

- **age** is not recorded

3.3 B

- correlation is not causation; a third factor, the weather, explains both

3.4

(a) any one, with a method:

- the **busiest times** of day: count journeys in each hour
- the **average speed** on each route: group by location and average the speeds
- where buses are **delayed**: find locations where speed is often low

(b)

- **remove** or **flag** them (clean the data) before the analysis
- they are **impossible** values (errors), and would make averages and conclusions wrong

(c)

- people **without** smartphones (e.g. many older or poorer passengers) are **not represented**

- so the data may not show the journeys or needs of **all** passengers, and decisions may favour one group