

# Exercise walk-through

## Extracting Information from Data ■ 2.3

eXercise

## You must be able to

- explain how programs find **patterns** 模式和 **trends** 趋势 in large data sets
- use **filtering** and **cleaning** 清洗 to prepare data
- distinguish **correlation** 相关 from **causation** 因果
- describe how **metadata** 元数据 and a **visualization** 可视化 help

## Method — Patterns and cleaning

Programs process large data sets to reveal **patterns and trends**. Raw data is first **filtered** and **cleaned** (removing errors and duplicates).

## Method — Correlation vs causation

Two things moving together is **correlation**; it does **not** prove one **causes** the other — a hidden factor may drive both.

## Method — Metadata and visualization

**Metadata** is data about data (e.g. a photo's date and location). A **visualization** (chart or graph) communicates insight. How data is **collected** can introduce bias.

## Practice 2.1 ■ 2 marks

State the difference between correlation and causation.

## Solution 2.1

Method: Correlation vs causation

### Worked steps

correlation is two things changing together; causation is one thing actually making the other happen **B1** **B1**.

## Practice 2.2 ■ 1 mark

Define metadata.

## Solution 2.2

### Worked steps

data about data (e.g. the date, size, or location of a file) **B1**.

## Practice 2.3 ■ 1 mark

State one reason to clean data before analysing it.

## Solution 2.3

### Worked steps

to remove errors, duplicates, or missing values so the analysis is reliable **B1**.

## Exam-style 3.1 ■ 1 mark

Data about data, such as a photo's date and location, is called

**A** metadata

**B** a variable

**C** a byte

**D** a pixel

## Solution 3.1

Method: Metadata and visualization

**A** metadata



**B** a variable

**C** a byte

**D** a pixel

### Worked steps

**A.**

## Exam-style 3.2 ■ 1 mark

Two variables that rise together show

**A** causation

**B** correlation

**C** an error

**D** nothing at all

## Solution 3.2

**A** causation

**B** correlation 

**C** an error

**D** nothing at all

### Worked steps

**B.**

## Exam-style 3.3 ■ 1 mark

A chart shows that ice-cream sales and drowning incidents both rise in summer.

- (a) Name the relationship the chart shows.
- (b) State whether one directly causes the other.
- (c) Name a likely hidden factor behind both.

## Solution 3.3

Method: Correlation vs causation

### Worked steps

(a) correlation **B1**. (b) no **B1**. (c) hot weather (more people swim and buy ice cream) **B1**.