

Exercise walk-through

2.3

Extracting Information from Data

AP Computer Science Principles

Extracting information from data

Programs turn raw **data sets** into **information**, finding **patterns** and **trends**.

Filtering

keep the rows you want

Cleaning

fix duplicates, blanks, errors

Beware: a **correlation** never proves **causation**, and **bias** in how data was collected can mislead. **Meta-data** & **visualization** help.



Hard-disk platters and head: data compressed and stored as magnetic patterns

From data to information

**data
information**

into infor-

data becomes useful when patterns, trends and answers are extracted from it

filtering

keeping only the rows that meet a condition

cleaning

removing errors, duplicates and incomplete records before analysing

visualizing

a chart makes a pattern visible that a table hides

the caution

large data sets reveal **correlation**, which is not causation – and they raise **privacy** questions about the people in them

Programs process data at scales no human can. That is the benefit *and* the reason the privacy question is on this paper.

Exercises 2.1-2.3

2.1 State the difference between correlation and causation.

[2]

2.2 Define metadata.

[1]

2.3 State one reason to clean data before analysing it.

[1]

Exercises 2.1-2.3

2.1 State the difference between correlation and causation.

[2]

2.1

- correlation is two things changing together; causation is one thing actually making the other happen

2.2 Define metadata.

[1]

2.2

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

2.3 State one reason to clean data before analysing it.

[1]

2.3

- to remove errors, duplicates, or missing values so the analysis is reliable

Exercises 3.1-3.2

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

[1]

☐

A metadata

B a variable

C a byte

D a pixel

3.2 Two variables that rise together show

[1]

☐

A causation

B correlation

C an error

D nothing at all

Exercises 3.1-3.2

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

[1]

☐

A metadata

B a variable

C a byte

D a pixel

3.1 A

- metadata.

3.2 Two variables that rise together show

[1]

☐

A causation

B correlation

C an error

D nothing at all

3.2 B

- correlation.

Exercise 3.3 (a)

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

(a) Name the relationship the chart shows.

[1]

Exercise 3.3 (a)

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

(a) Name the relationship the chart shows.

[1]

(a)

- correlation

Exercise 3.3 (b)

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

(b) State whether one directly causes the other.

[1]

Exercise 3.3 (b)

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

(b) State whether one directly causes the other.

[1]

(b)

- no

Exercise 3.3 (c)

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

(c) Name a likely hidden factor behind both.

[1]

Exercise 3.3 (c)

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

(c) Name a likely hidden factor behind both.

[1]

(c)

- hot weather (more people swim and buy ice cream)