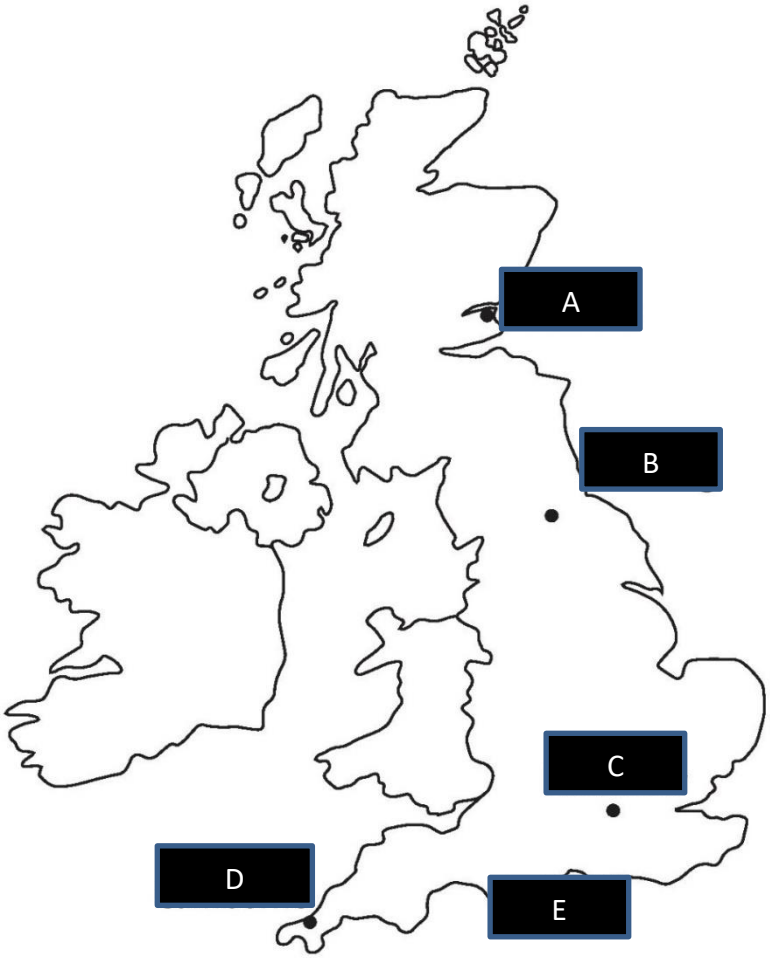
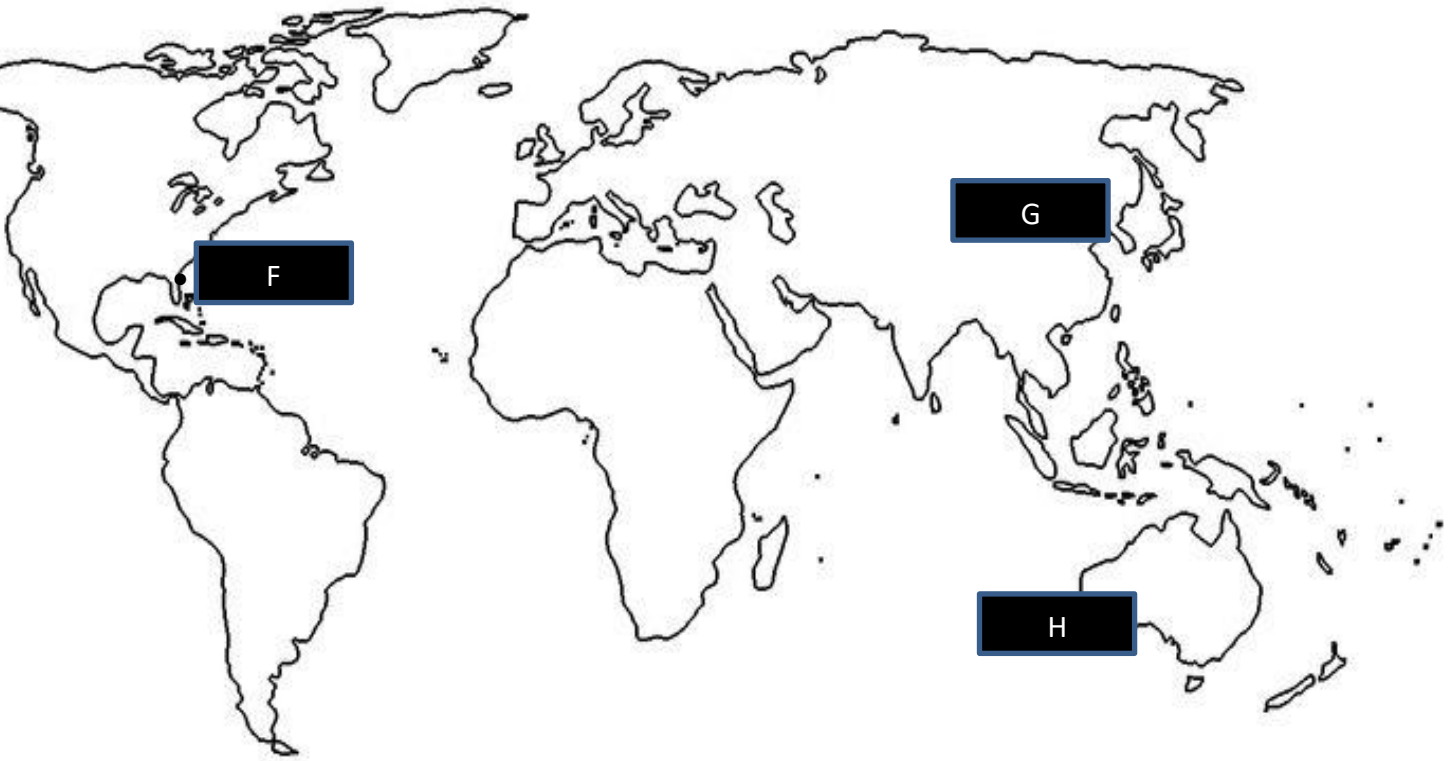


1. i) Name the towns at A, B, C, D and E



ii) a) Name the cities at F, G, and H
b) In which countries are cities F, G, and H ?

2. Sara is investigating the variation in the daily maximum gust, t kn, for Camborne in June and July 1987. She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.
(a) State the sampling technique used by Sara.
(b) From your knowledge of the large data set explain why this process may not generate a sample of size 20.
The data Sara collected are summarised as follows $n = 20, \sum t = 374, \sum t^2 = 7600$
(c) Calculate the standard deviation. Give your answer to 3 s.f.

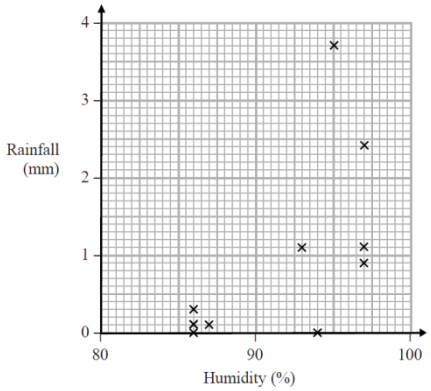
3. Sara was studying the relationship between rainfall, *r* mm, and humidity, *h*%, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set. She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1, \quad Q_2 = 0.9, \quad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.
(a) Show that *r* = 20.6 is an outlier.
(b) Give a reason why Sara might:
(i) include this day's reading (ii) exclude this day's reading.
Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of *r* and *h*.



c) Give an interpretation of the correlation between rainfall and humidity.
The equation of the regression line of *r* on *h* for these 10 days is *r* = −12.8 + 0.15*h*
(d) Give an interpretation of the gradient of this regression line.
(e) (i) Comment on the suitability of Sara's sampling method for this study.
(ii) Suggest how Sara could make better use of the large data set for her study.

4. Lauren wants to find the average daily mean windspeed in Hurn in 1987.
She only has access to the large data set. She uses it to obtain a random sample of the daily mean windspeeds, *t* knots, on *n* days in Hurn in 1987.
The data collected by Lauren are summarised as follows

$$\sum (t - 5) = 55, \quad \overline{t} = 10$$

(a) Find *n*.
Lauren uses the same sampling method to estimate that the average daily mean windspeed in Hurn in 2015 was 11 mph.
(b) Convert 11 mph into knots.
(c) Hence, compare the average daily mean windspeed in Hurn in 1987 and 2015.
(d) With reference to the large data set, state **one** limitation of your conclusion in part (c).
(e) Explain how Lauren can
(i) improve her data collection method
(ii) improve her data processing
to allow for a more reliable comparison in part (c).

Answers

1. i) A = Leucahrs, B = Leeming, C = Heathrow, D = Camborne, E = Hurn
ii) a) E = Jacksonville, F = Beijing, G= Perth
b) E = USA, F= China, G= Australia

2. a) Systematic
b) In the Large Data Set some days have gaps because the data was not recorded.
c) 5.51

3. b) i) e.g. It is a piece of data and we should consider all the data
ii) e.g. It is an extreme value and could unduly influence the analysis **or**
It could be a mistake.
c) e.g. "as humidity increases rainfall increases"
d) e.g. a 10% increase in humidity gives rise to a 1.5 mm increase in rainfall **or**
represents 0.15mm of rainfall per percentage of humidity
e) i) Not a good method since only uses 11 days from one location in one month
ii) e.g. She should use data from more of the UK locations and more of the months **or**
using a spreadsheet or computer package she could use all of the available UK data

4. a) 11
b) 9.56 knots
c) Hurn had a higher average daily mean windspeed in 1987 than in 2015
d) The Large Data Set only covers the months May to October **or**
The Large Data Set does not cover the whole year
e) i) use a larger data set so it is more representative
ii) consider standard deviation/variation, so she can compare/take into account the spread of the data **or**
consider another average/the median/the mode, so she can compare with respect to other averages **or**
consider/exclude outliers, to avoid the average being influenced by extreme values