14 The pre-release material contains information concerning the median income of taxpayers in £ and the percentage of all pupils at the end of KS4 achieving 5 or more GCSEs at grade A*–C, including English and Maths, for different areas of London.

Some of the data for 2014/15 is shown in **Fig. 14.1**.

Fig. 14.1

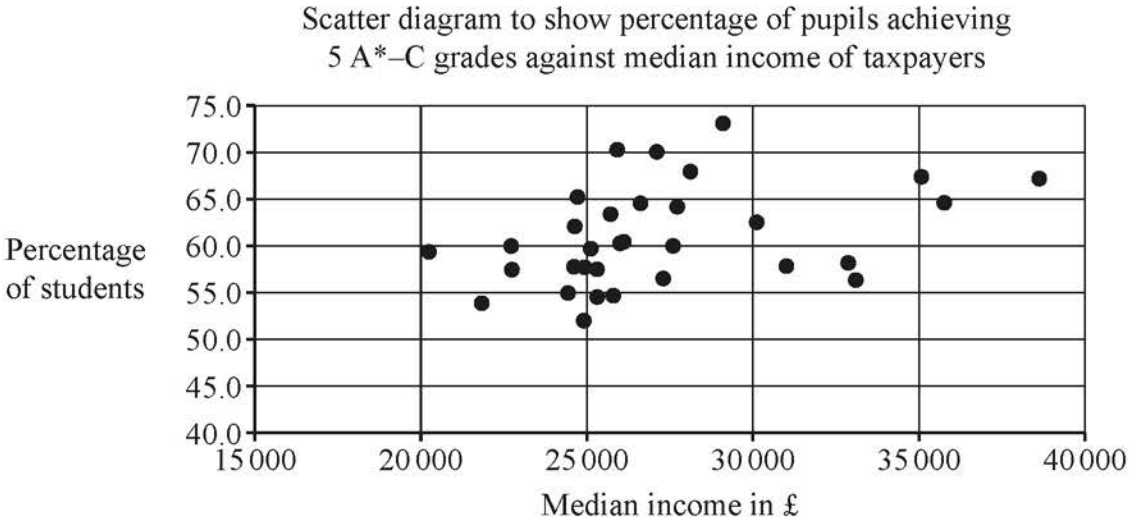
	Median Income of Taxpayers in £	Percentage of Pupils Achieving 5 or more A*–C, including English and Maths
City of London	61 100	#N/A
Barking and Dagenham	21 800	54.0
Barnet	27 100	70.1
Bexley	24 400	55.0
Brent	22 700	60.0
Bromley	28 100	68.0

A student investigated whether there is any relationship between median income of taxpayers and percentage of pupils achieving 5 or more GCSEs at grade A*–C, including English and Maths.

(a) With reference to **Fig. 14.1**, explain how the data should be cleaned before any analysis can take place. **[1]**

After the data was cleaned, the student used software to draw the scatter diagram shown in **Fig. 14.2**.

Fig. 14.2



The student calculated that the product moment correlation coefficient for these data is 0.3743.

(b) Give **two** reasons why it may not be appropriate to use a linear model for the relationship between median income of taxpayers in £ and the percentage of all pupils at the end of KS4 achieving 5 or more GCSEs at grade A*–C. **[2]**

The student carried out some further analysis. The results are shown in **Fig. 14.3**.

Fig. 14.3

	median income of taxpayers in £	percentage of pupils achieving 5+ A*–C
mean	27216	61.0
standard deviation	4177.5	5.32

The student identified **three** outliers in total.

(c)

- Use the information in **Fig. 14.3** to determine the range of values of the median income of taxpayers in £ which are outliers.
- Use the information in **Fig. 14.3** to determine the range of values of the percentage of all pupils at the end of KS4 achieving 5 or more GCSEs at grade A*–C which are outliers.
- On the copy of **Fig. 14.2** in the **Printed Answer Booklet**, circle the **three** outliers identified by the student. **[4]**

The student decided to remove these outliers and recalculate the product moment correlation coefficient.

(d) Explain whether the new value of the product moment correlation coefficient would be between 0.3743 and 1 or between 0 and 0.3743. **[1]**