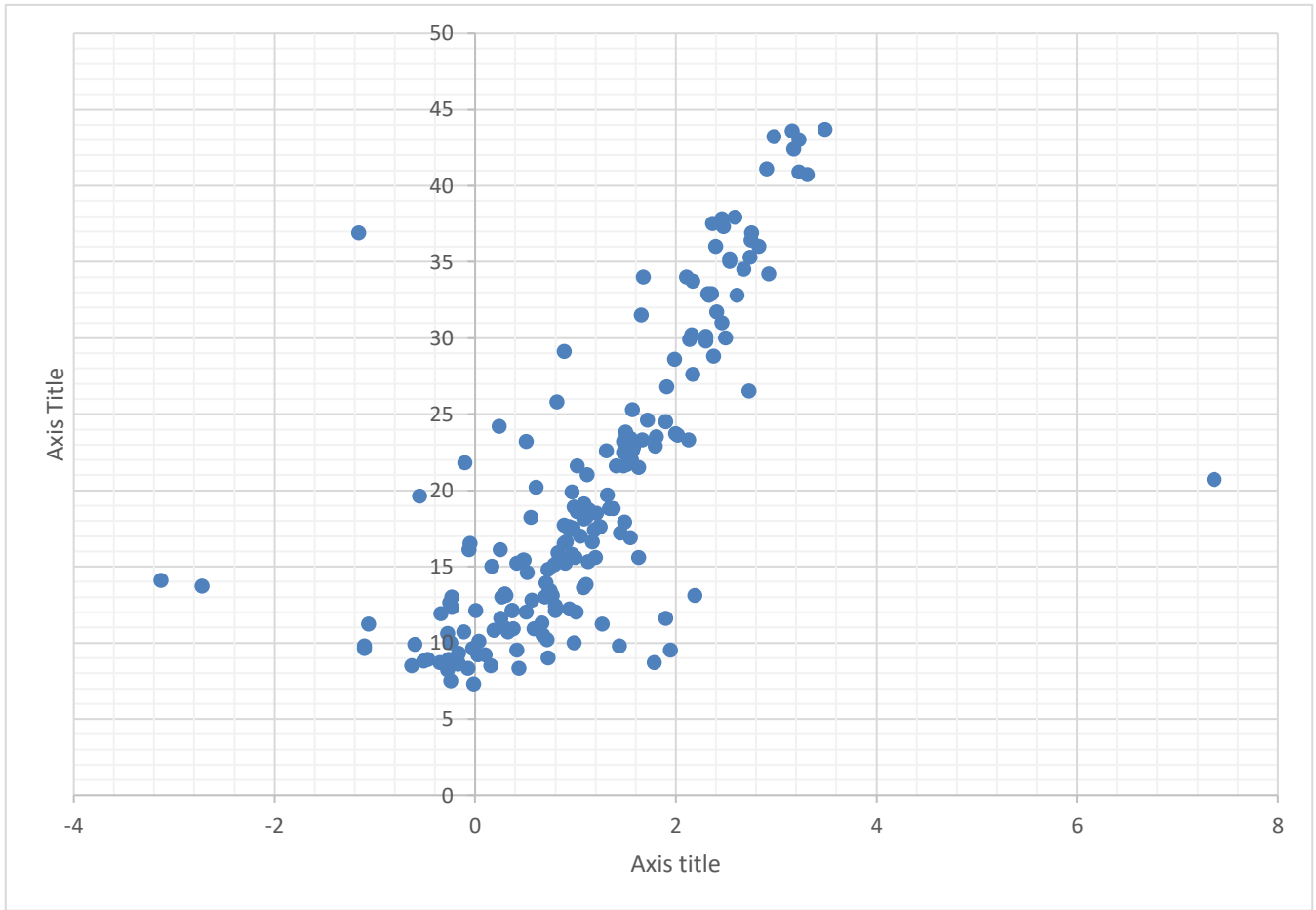


Correlation

The [CIA World Factbook](#) contains information on many countries.

The scatter diagram below a plot of data from 188 countries.



One axis title should be **Birth rate** (*births per 1000 population*) and the other should be **% Population growth**.

1. Work out which axis title is which and complete the statement below:

I think horizontal axis represents because

This means the vertical axis must be

2. Describe the general trend shown in the data.

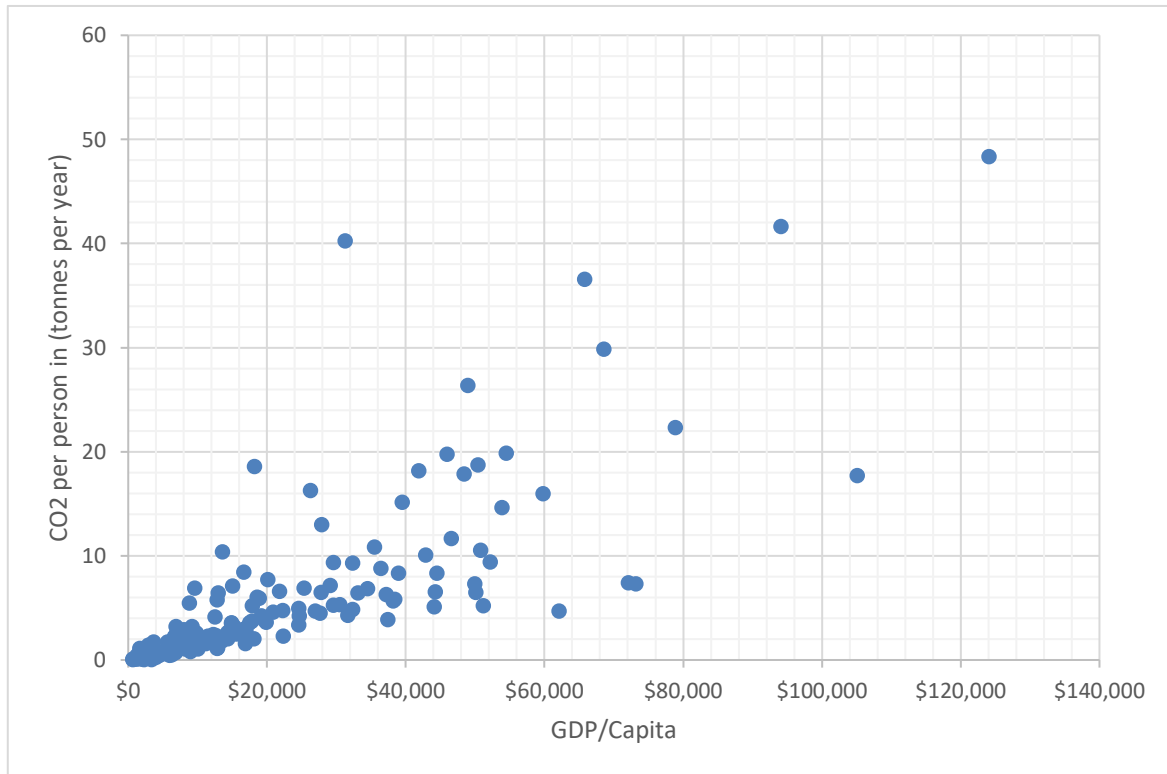
Outliers are points which don't seem to fit the general pattern.

3. Circle any points which you think don't seem to fit the general pattern.
4. Find the name of the countries you identified as outliers. (Use the spreadsheet or the data sheet)
5. Find out about these countries. Can you find any reason why they may not follow the general pattern?

(The population change which appears in the CIA World Factbook is calculated by taking (births – deaths) + (people entering – people leaving) then dividing this by the population, this fraction is then expressed as a percentage.)

Data from CIA world Factbook <https://www.cia.gov/library/publications/resources/the-world-factbook/>

Here is another scatter diagram constructed using data on 180 countries from the CIA World Factbook.



Gross domestic product (GDP) is the total earnings of a country. Per capita means per person. So GDP per capita, is the earnings of the country divided by the population of the country.

6. Explain why GDP per capita may be better measure of wealth than GDP.

7. What general trend does the data show?

8. Can you suggest a possible reason for this trend?

Below are two scatter diagrams. A line of best fit has been added to each of the scatter diagrams.

Diagram A

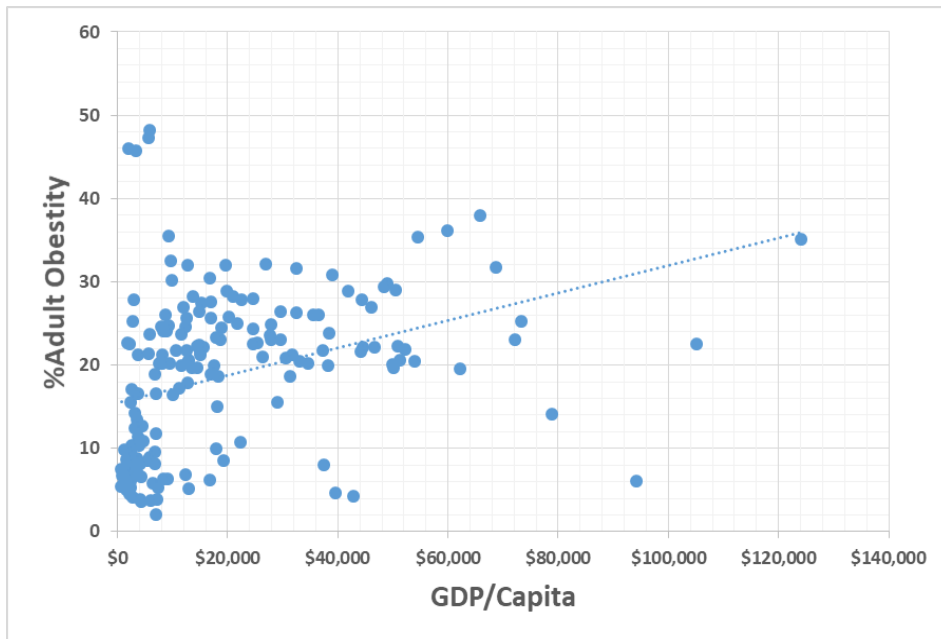
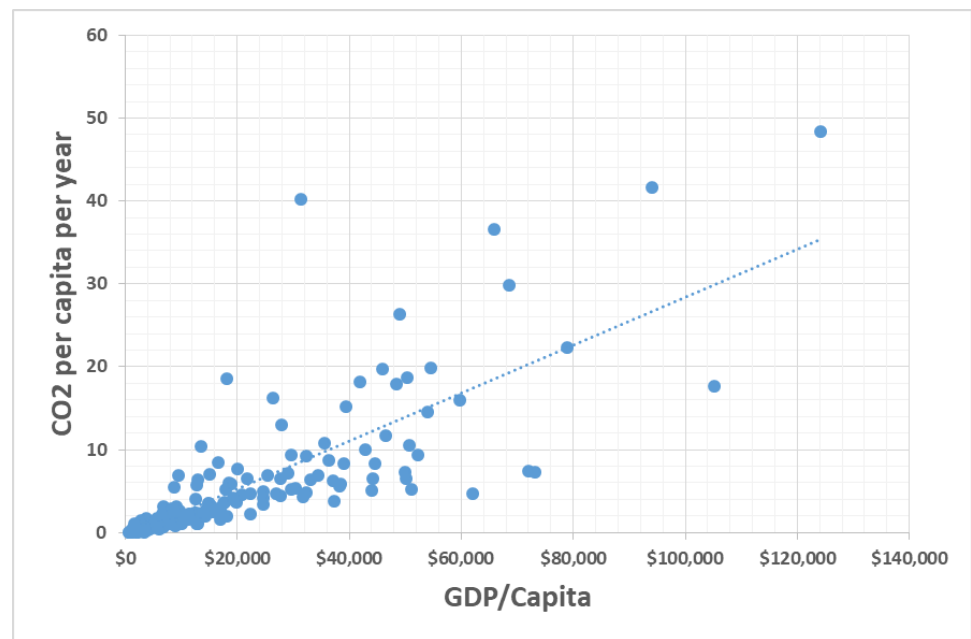


Diagram B



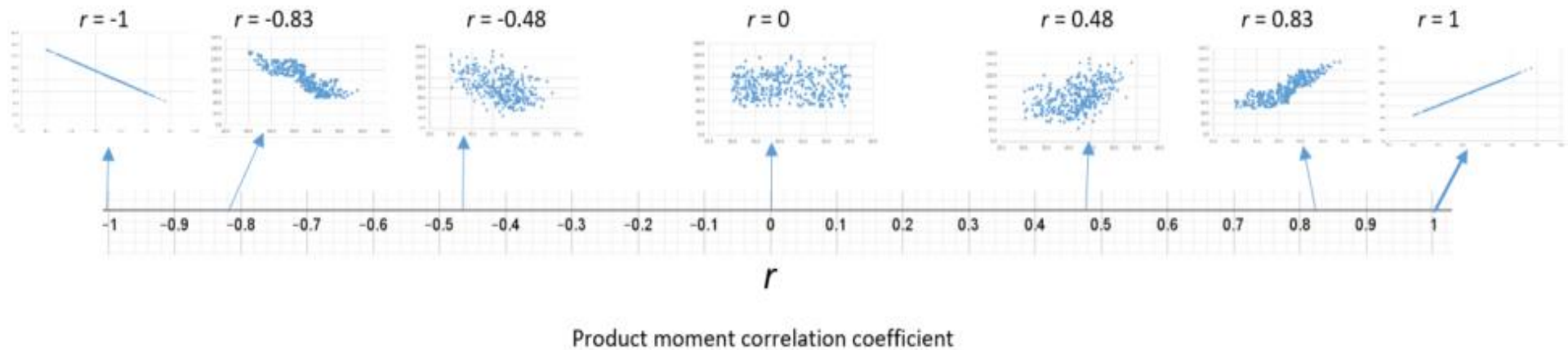
9. Which of the lines appears to fit the data better, the one in diagram A, or the one in diagram B?
(Explain your reasoning)

Using (r) to compare the strength of correlations

The product moment correlation coefficient (r) is used to compare the strength of the correlation between two variables. Here are some diagrams with different r values. The scales on the x and y axis are the same in each diagram to make it easier to see the change in the shape of the scatter directly, without having to think about the effect of the scale.

An r value of -1 represents a perfect negative correlation, whilst an r value of 1 is a perfect positive correlation and $r = 0$ when there is no correlation.

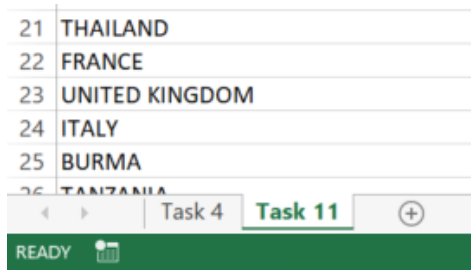
Look at the diagram below



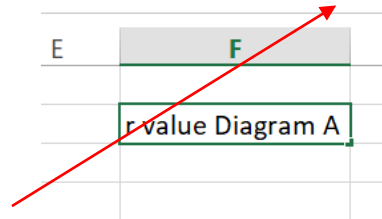
10. You need the spreadsheet “correlation tasks”.
Follow the instructions on the next page to calculate the r value for both diagrams.

Using the =correl command to calculate r

1. Select the tab marked task 11

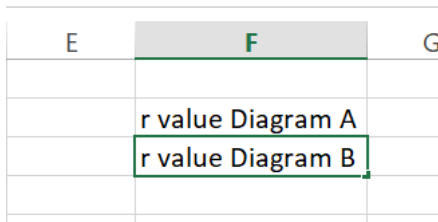


2. Select cell F2 type in r value Diagram A

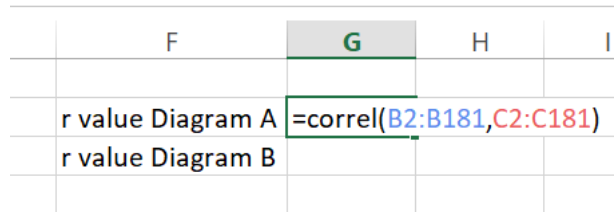


Double click here to resize the column width

3. Select cell F3 type in r value Diagram B

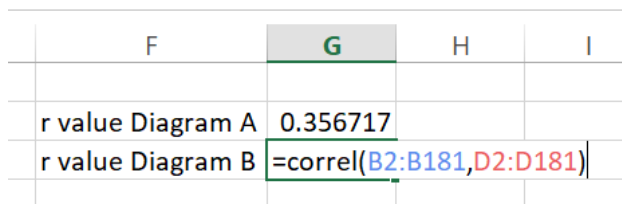


4. Select cell G2 and enter the formula
=CORREL(B2:B181,C2:C181)



Press enter, and the value 0.356717 should appear in G2

5. Select cell G3 and enter the formula
=CORREL(B2:B181,D2:D181)



Press enter, and the value 0.795267 should appear in G3

6. You have just used Excel to calculate the Product Moment Correlation Coefficient (**r**) for both lines.

In Diagram A $r = 0.356717$

In Diagram B $r = 0.795267$

11. Try the correlation card matching activity.