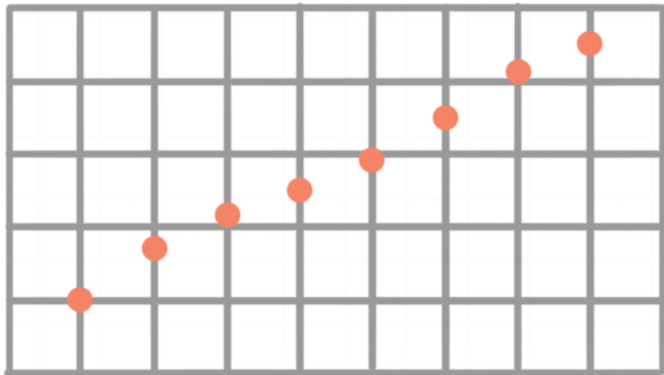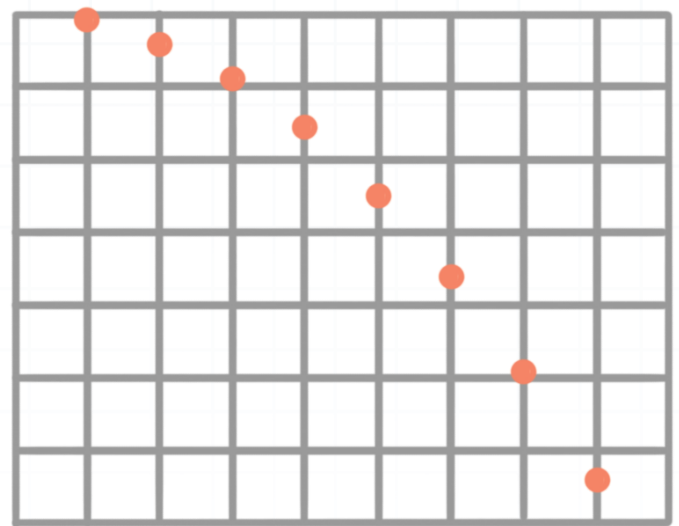**Topic**: Correlation coefficient and the residual

Question: Which scatterplot has a correlation coefficient closest to -1?
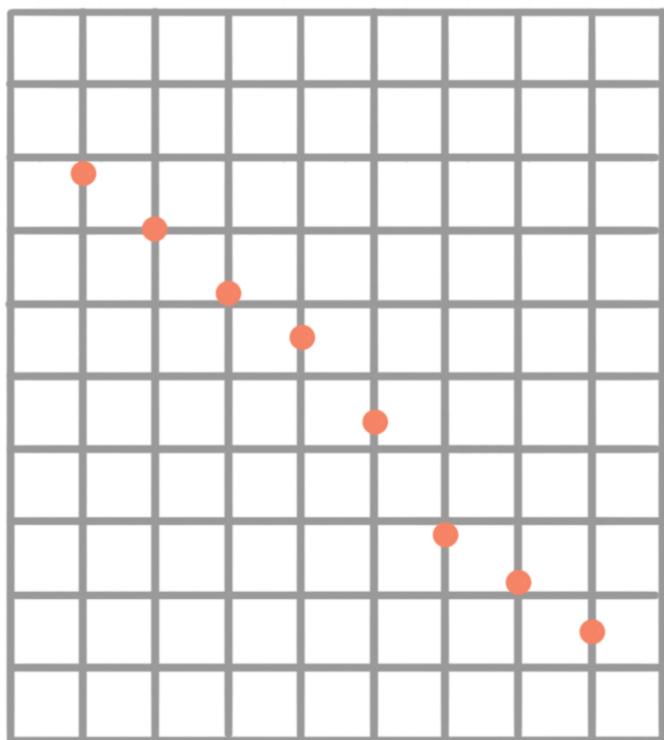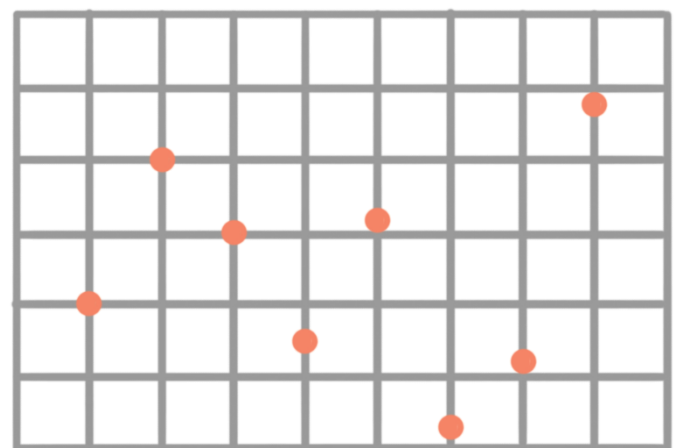
Answer choices:

A

B

C

D

Solution: C

A correlation coefficient of -1 means the trend line has a negative slope and the points are very close to forming a line.

Answer choice A is an example of an R-value close to $+1$ because the points rise from left to right and are close to forming a line.

Answer choice B is an example of points that have a quadratic relationship. This might be estimated with a trend line and give a negative R-value, but there's a better choice.

Answer choice C has points that are more linear than choice B and they decrease from left to right, so this is the best answer.

Answer choice D has points that are spread out, so the correlation coefficient is likely very close to 0.

Topic: Correlation coefficient and the residual

Question: The line of best fit for a data set is given by $y = 2.0929x - 0.4429$. What's the value for the residual of the point $(6,12)$?

Answer choices:

A 12.1145

B -12.1145

C 0.1145

D -0.1145

Solution: D

The residual of a point is its actual value minus its predicted value.

We can calculate the actual value by plugging in the x-coordinate of the point $(6,12)$, into the line of best fit. The value $x = 6$ gives

$$\hat{y} = 2.0929(6) - 0.4429$$

$$\hat{y} = 12.1145$$

This is the predicted value. The actual value is the y-value of the point $(6,12)$. The residual is

actual $-$ predicted

$12 - 12.1145$

0.1145

Topic: Correlation coefficient and the residual

Question: Which statement is false?

Answer choices:

A   An R-value close to 1 or -1 means the line of best fit is perfect at making predictions from your data, you do not need any other checks.

B   Always draw a scatterplot of your data points first before you perform a linear regression to get an idea of the shape of the data.

C   When you check the residual plot, it needs to be in a random pattern with points above and below 0 in order for the line of best fit to be a good predictor of the data.

D   The value of a data point is always equal to the value from the linear model, plus the residual value.

Solution: A

An R-value close to 1 or -1 could nicely describe the data that you got, but you need to be careful here.

Sometimes the R-value can be close to 1, but the data is just a small snapshot of another type of function. That's why it's important to both draw the scatter plot, as well as to calculate the residuals and look at the residual plot. These graphs, especially the graph of the residuals, can show you what the R-value can't.

For instance, the graph could tell you if your data is doing something unexpected, or if it's a small section of an exponential or parabolic function.