The following are descriptions for the scatterplots on the correlation quiz.

Graph 1:
A group of points is in the lower left corner with two points way in the upper right corner of the plot.

The issue here is outliers. The two outliers are increasing what the correlation would have been without the two data points. This could also be a problem of range restriction in that there are no x-axis values between the main group of points and the outliers. It is difficult to tell in this case.

Graph 2:
The points form a bowl type of shape.

Nonlinearity is the issue with this graph. Because the points fall in a parabolic pattern instead of a linear pattern, the correlation will underestimate the true relationship between the two variables.

Graph 3:
There is a circular glob of points. There appears to be no correlation. The x-axis ranges from 0 to 1 instead of from -3 to 3 as in many of the other graphs.

We have a case of restriction of range. There is not much variation in X. Thus, it is difficult to find a relationship between X and Y.

Graph 4:
There is a group of points with a negative looking trend. Some of the points are the plus symbol while other points are stars. The stars tend to clump in the lower right, and the pluses tend to clump in the upper left.

This is the case of having heterogeneous subgroups. The relationship between X and Y is not very strong. Mixing non-homogeneous groups, however, makes the relationship between the variables appear stronger because of a mean difference between the groups.

Plot 5:
The scatterplot seems to show a negative trend. A large group of points lines up at the Y value of 1.5, and no points go any higher than 1.5. It looks like someone chopped off the top of the scatterplot.

This is once again our enemy of range restriction. The Y values seem to be stopped at a maximum by some sort of ceiling effect. The value of our correlation will be lower than it should be.

Plot 6:
A group of points going from the lower left to the upper right of the scatterplot seem to show a positive linear trend. There is one point in the extreme upper left of the graph.

These data have an outlier. In this case, the outlier lowers the value of the correlation.

Through these graphs, we have examples of four common problem cases when using the Pearson correlation coefficient: outliers, range restriction, heterogeneous subgroups, and non-linear effects. Although these graphs do not represent every case possible, they give you an idea of the problems.

Outliers and heterogeneous subgroups can have the effect of either decreasing or increasing the correlation. Range restriction and non-linear effects almost always decrease the correlation.