

# Understanding Correlation

## Correlation in the Visual Ranking Tool

A correlation between two sets of numbers is a measure of the degree that one can know about one set by knowing about the other. In the case of sorted lists, it tells you how closely your list compares with another person's list. Without seeing the other list, you can know a great deal about it just by considering the correlation.

The easy cases are correlations of 1.0 and -1.0. With a 1.0, you know the other list is exactly like yours. A correlation of -1.0 means that the order of other list is exactly opposite of yours. If you see a correlation of over .9, you know that the other list is essentially like yours, but a couple of items in the other list have moved one or two places up or down.

When the correlation is in the moderate range, the lists may look quite different. In the image below, the lines between like items seem to be criss-crossing in a random jumble, yet the correlation is high. To make sense of this apparent discrepancy, ask yourself "What do we agree on?" As you see, a list of the top five and bottom five choices would have the same items. So the students agree on what items are high and low and exact agreement could be negotiated by moving items a short distance.

Project Name: **Inventions that changed our lives**  
 Team ID: **Group1**  
 Prompt: **Rank these inventions in order of how they have improved peoples' lives.**

Compare your team with: **Group4**  
 There is a strong positive correlation of 0.64

Wheel	Light bulb
Printing press	Refrigeration
Refrigeration	Wheel
Light bulb	Gun powder
Gun powder	Printing press
World Wide Web	Clock
Clock	Personal computer
Personal computer	World Wide Web

What can you tell about another list if the correlation between your list and the other is close to 0? Not much. You know that the lists aren't the same and they aren't opposite, but other than that, you really can't predict what the order of the other list will be.

A correlation is derived by "measuring" the distance of the lines between each pair of items. When you see two lists that are the same, the lines between items are as short as they can be. When lists are the opposite the lines have the greatest total length possible. High correlations reflect minor disagreements which are represented by short diagonal lines. A very low correlation has relatively long lines between many of the items.

While interpreting the correlation values is relatively simple, not all values are possible for some lists. For example, if the list has only two items, then two students can only agree exactly (with a correlation of 1.0) or disagree completely (with a correlation of -1.0). As the list grows, there are more combinations of possible and, therefore, more possible correlations. The *Visual Ranking Tool* uses the following values to determine the phrases that describe the correlation value.

Correlation	Text that is Displayed in the Visual Ranking Tool
-1.0 to -.8	There is a very strong negative correlation
-.6 to -.79	There is a strong negative correlation
-.4 to -.59	There is a moderate negative correlation
-.2 to -.39	There is a weak negative correlation
-.01 to -.19	There is a very weak negative correlation
0 to .19	There is a very weak positive correlation
.2 to .39	There is a weak positive correlation
.4 to .59	There is a moderate positive correlation
.6 to .79	There is a strong positive correlation
.8 to 1.0	There is a very strong positive correlation

Part of interpreting the correlation value is deciding whether differences between lists are desirable or undesirable. This is a good job for your students. Since lists with a strong negative correlation often reflect different, equally-valid viewpoints, let your students know that the word "negative" in a negative correlation means "opposite" rather than "bad."

## Spearman Rank Order Correlation Coefficient

The correlation that is displayed in the *Visual Ranking Tool* is called the Spearman Rank Order Correlation coefficient, a measure that is reserved for looking at data sets that are arranged in rank order.

It can be applied to compare two independent random variables, each at several levels. Spearman's rank correlation works on ranked data, rather than directly on the data itself. Spearman's method works by assigning a rank to each observation in each group separately.

If the mathematics of the statistic interests you, the formula used in the *Visual Ranking Tool* is:

$$rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$