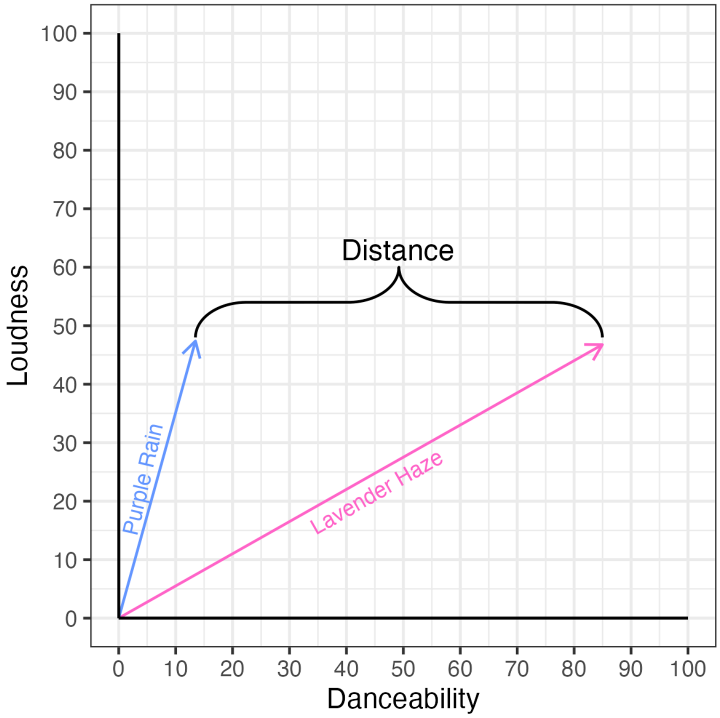
# The Mathematics of Euclidean Distance

In the previous set of activities, we computed the similarity between songs using Euclidean Distance. Mathematically, we define the distance between two observations *a* and *b* as:

$$d(a, b) = \sqrt{\sum_{i=1}^{p} \left(a_i - b_i\right)^2}$$

where, *p* is the number of measured attributes for the observations. In higher levels of mathematics, $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are vectors in *p*-space. To help you think about this, we will simplify this to two-space, something you graph in a lot in high school mathematics. Being in two-space means that there are two attributes measured for each song (e.g., danceability and loudness). You can visualize a vector as an arrow that has the butt-end of the arrow at the origin (0, 0), and the arrow-end at the (danceability, loudness)-value for the song in question.

For example, consider the songs *Lavender Haze* (danceability = 84.96, loudness = 46.76) and *Purple Rain* (danceability = 13.48, loudness = 47.39). Each song would constitute a different vector (arrow) and the Euclidean Distance is visualized as the distance between the arrow tips.



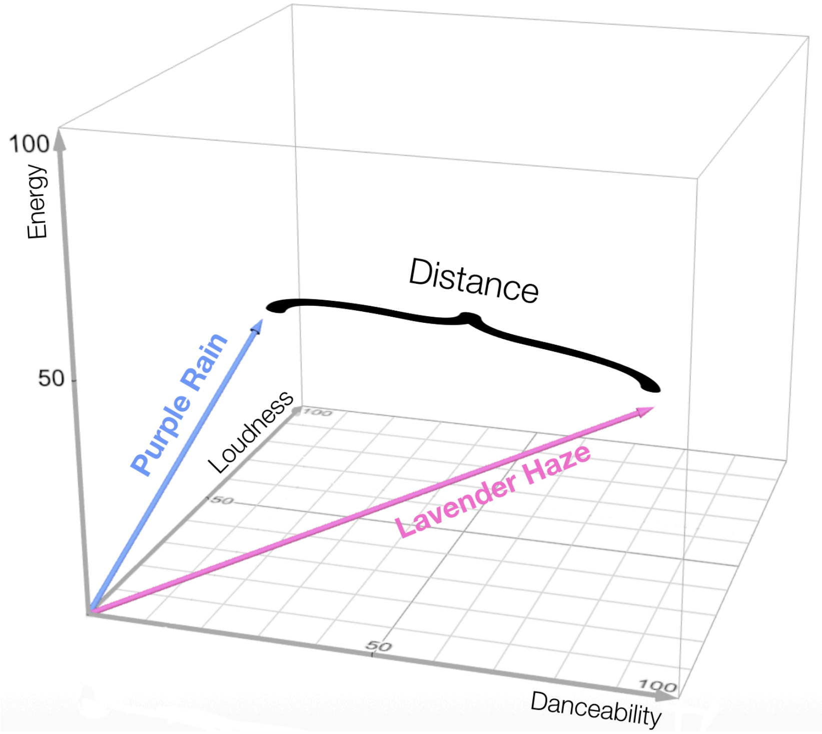
Songs that were more similar to each other had smaller Euclidean Distance measures, while songs that were less similar to each other had larger Euclidean Distance measures.[1]

---

[1] Because larger distances mean that the songs are less similar, sometimes data scientists refer to Euclidean Distance as a *dissimilarity* measure.

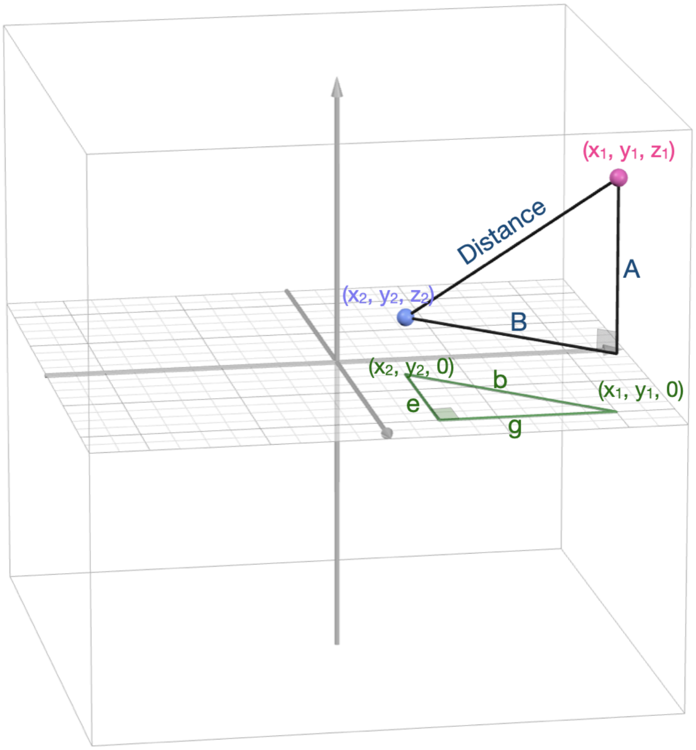
## Euclidean Distance in More Than Two Dimensions

We can also extend the idea of Euclidean Distance to measure similarity when we have more than two attributes. (For example, imagine we had measured three metrics for our songs: loudness, danceability, and energy.) Again, we are measuring the distance between two vectors, however this time they are plotted in 3-space; there would be a loudness axis, a danceability axis, and an energy axis.



Essentially we need to find the distance between two points in 3-space. Let's say that these two points have the coordinates ($x_1$, $y_1$, $z_1$) and ($x_2$, $y_2$, $z_2$), respectively. The following plot shows two such points. To determine the distance between the two points, we can form a right triangle such that the hypotenuse of that triangle extends between the two points and one of the legs is parallel to the vertical axis (see black triangle in figure below).

We need to find the length of both legs of this right triangle, namely A and B, so that we can use the Pythagorean Theorem to find the distance between the points ($x_1$, $y_1$, $z_1$) and ($x_2$, $y_2$, $z_2$). Because we created this triangle so that A was parallel to the vertical axis, the length of A is simply ($z_1 - z_2$). To find the length of B, we project that leg onto the plane created by the non-vertical axes. (This projection creates the line segment b, which is now in two-space.) This segment is the hypotenuse of another right triangle (green triangle in figure below) that connects the points $(x_1, y_1)$ and $(x_2, y_2)$.



1. Find the length of b in the green triangle.

2. Since b is the projection of B, both b and B have the same length. Now that you know the length of B and A, use those values to determine the hypotenuse of the black triangle, which is the distance between $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$.

Euclidean Distance between two points in 3-space is just an extension of the Euclidean Distance between two points in 2-space! We are computing the squared differences between the corresponding coordinates, summing them, and finding the square root of that sum. In fact, if we move on to 4-, 5-, or even n-space, the Euclidean Distance is found similarly, we just have additional coordinates.

$$Distance = \sqrt{\sum_{i=1}^{n} (Dimension\ i_1 - Dimension\ i_2)^2}$$

As an example, to find the Euclidean Distance between *Lavendar Haze* (danceability = 84.96, loudness = 46.76, energy = 35.18) and *Purple Rain* (danceability = 13.48, loudness = 47.39, energy = 45.43), we compute:

$$Distance = \sqrt{(84.96 - 13.48)^2 + (46.76 - 47.39)^2 + (35.18 - 45.43)^2} = 72.21$$