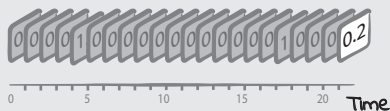


Ever heard about a **time curve**?

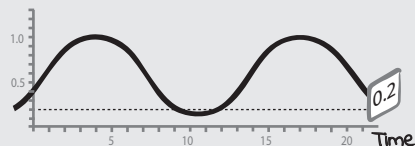
Um, no, tell me more please.



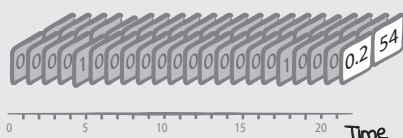
Sure! Line graphs are only really useful for single scalar-value variables, like bacterium length.



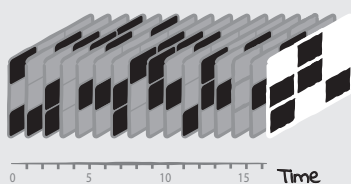
They work, as you know, by representing the value as height.



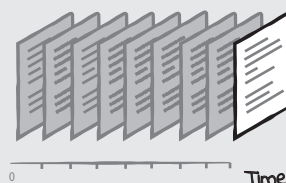
But how do we deal with more complex variables, like those composed of several values?



Or video footage where each pixel in every frame is a value of its own?



Or wikipedia articles, where the variable is an article being lengthened, shortened and revised over time?



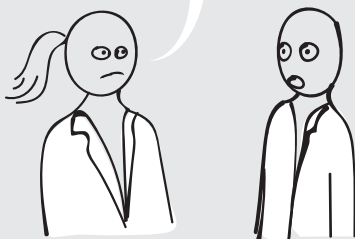
To be able to map any change to such complex variables, we can compare it at two different points in time.



Edit distance: 1789

For example, the difference between a text at point A and B could be measured by their "edit distance".

And DNA? where does bacterial DNA enter the picture?

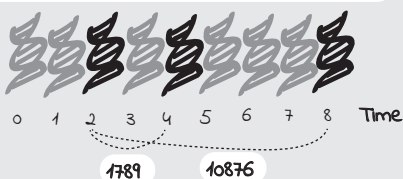


Allow me some dramatic build-up, will you?

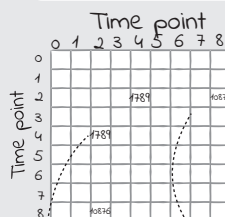


Anyway - DNA is ultimately just a text string right? Like an evolving wikipedia-article.

Now suppose that we calculated the edit distance between each point and every other point, to see how the DNA at one point differs from DNA at another.



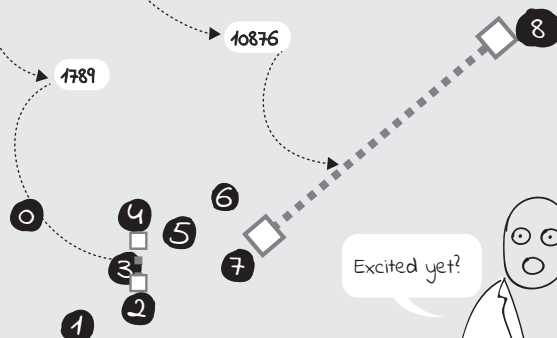
Now for the grand reveal.



Having filled out the entire distance matrix we know the edit distance, the difference, between any two points.

For example, point 8 is about 6 times more different to point 2 than point 4.

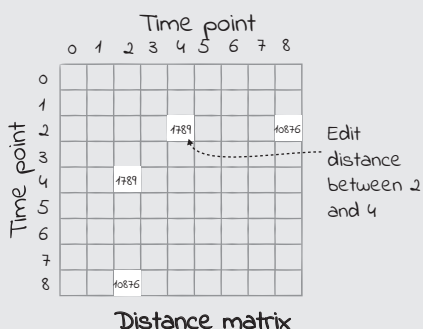
Now what if we translated that to **literal** distances between time points?



Excited yet?



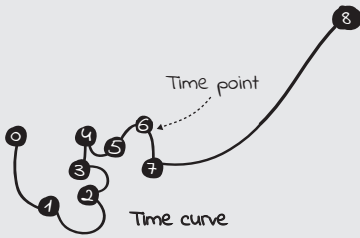
That's a whole lot of edit distances, which we'd better store in a "distance matrix".



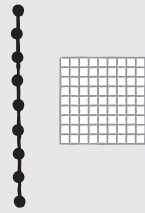
Edit distance between 2 and 4

Distance matrix

Proceeding to link these time points together in their time order gives us a curled up curve, called a **time curve**.



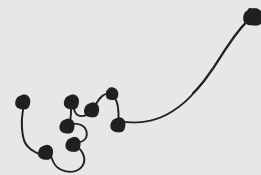
The process of turning a distance matrix into a time curve is in reality a complex optimisation problem.



>

Black Box

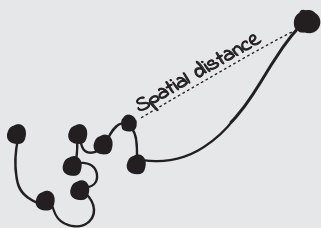
>



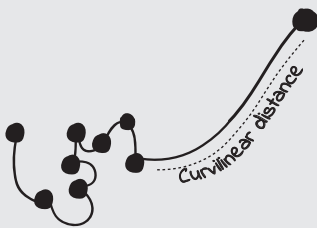
It uses an algorithmic approach called **multi-dimensional scaling** that exists in many different variants.



once the time curve has been produced, there are three different kinds of information you can get from it. The most important being **spatial distance**, which gives the similarity between time points.



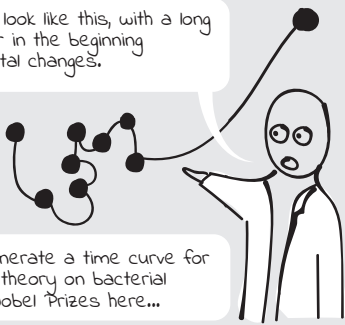
There is also the **curvilinear distance** between two time points, which approximately represents the total amount of change in the segment.



But on a higher level, structures emerge from the time curve that inform us of the underlying evolution processes. Things like...

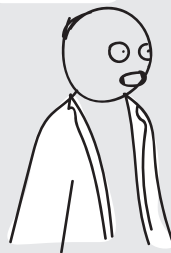


So an evolutionary spurt would look like this, with a long transition... whereas this cluster in the beginning reflects more gradual, incremental changes.



Now, Anna, could you please generate a time curve for my bacterial data to prove my theory on bacterial evolution spurts? we're talking Nobel Prizes here...

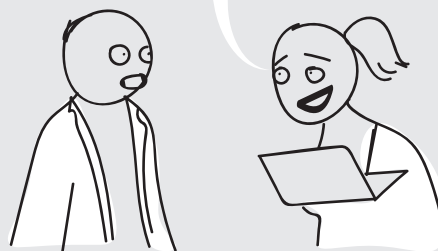
Anna?



okay so the bad news is that I miiiight have accidentally polluted your Petri dish this morning and the evolution you speak of never happened...



As for the good news...



... I did make you a time curve though!



* Learn more about the algorithm:
Bach, B., Congei, S., Hewlet, N., Madhyastha, T., Grabowski, T. & Dragicevic, P. (2016, January). Time Curves: Folding Time to Visualize Patterns of Temporal Evolution in Data. In IEEE Transactions on Visualization and Computer Graphics (Vol. 22, No. 1, pp. 4-7).