

What even is data?

July 25, 2025

1 Let's Talk Data:

We hear and talk a lot about *data* science or about large *data*. But what really are *data*? What does it look like? Does it have a structure we can envision? Think about the problem of trying to predict the temperature of a certain point in the city on the following day. We know (somehow) that the temperature will depend on the temperature of that location the previous day (T_{-1} , the humidity (in percentage) (H_{-1}) the amount of rainfall (in mm) (R_{-1}) and whether the point is located within the city or outside ($loc \in \{0, 1\}, \{in, out\}$). The datapoint then in this case looks like the observed values for the tuple $(T_{-1}, H_{-1}, R_{-1}, loc)$.