Let me explain MLE while we wait in line. As is often the case in statistics, our objective is to find an estimate for a model parameter using our data. In short, MLE is a method to estimate the parameter. The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed. The goal of MLE is to find the optimum wat to fit distribution to our data.

There are lots of types of distribution for different types of data. For example, normal, exponential and Gamma. We first have to decide which model we think best describes the process of generating the data. This part is very important. At the very least, we should have a good idea about which model to use. Usually we'll assume that the data generation process can be adequately described normal distribution. That means we expect most of the things. First we expect most of the measurements to be close to the mean or average.Also expect the measurements to be relatively symmetrical around the mean although the measurements are not perfectly symmetrical around the mean they are not crazy skewed to one side either this is pretty good normal distributions.Maximum likelihood estimation is a method that will find the values of μ and σ that result in the curve that best fits the data. Unfortunately most of the values we measured are far from the distributions average according to a normal distribution with a mean value over here the probability or likelihood of observing all these weights is low. What if we shifted the normal distribution over, so that its mean was the same as the average weight according to a normal distribution with a mean value. Here the probability or likelihood of observing these weights is relatively high. If we kept shifting the normal distribution over then the probability or likelihood of observing these measurements would go down. Again we can plot the likelihood of observing the data over the location of the center of the distribution.

Now we have to figure out the maximum likelihood estimate for the standard deviation. Again we can plot the likelihood of observing the data over different values for the standard deviation. Now we found the standard deviation that maximizes the likelihood of observing the weights. We measured this is the normal distribution that has been fit to the data by using the maximum likelihood estimations for the mean and the standard deviation. Now when someone says that they have the maximum likelihood estimates for the mean or the standard deviation or for something else you know, they found the value for the mean or the standard deviation or for whatever that maximizes the likelihood that you observed the things that you observed terminology alert in everyday conversation probability and likelihood mean the same thing. However in stats land, likelihood specifically refers to this situation we've covered here where you are trying to find the optimal value for the mean or standard deviation for a distribution, given a bunch of observed measurements this is how we fit a distribution to data.