The Matlab command \texttt{prctile(data,75)} computes sample percentiles in a way that is equivalent to the following procedure.

To get the $q^{th}$ quantile, or $100 \times q^{th}$ percentile, on data $x_1, x_2, \ldots, x_n$:

1. Order the data $x_(1) \leq x_(2) \leq \ldots \leq x_(n)$ and set $x_(0) = x_(1), x_(n+1) = x_(n)$, the min and max respectively.

2. Find $q \times n + .5$, and write it as an integer plus a decimal remainder $r \in [0, 1)$: $q \times n + .5 = i + r$.

3. The answer is the linear interpolation $x_(i) + r \times (x_(i+1) - x_(i))$.

Note that if $r = 0$, then $q = (i - .5)/n$, and the procedure says that the $q^{th}$ sample quantile is the $i^{th}$ data point.

Also, observe that the $q^{th}$ sample quantile is nearly the value $x$ where the empirical cdf ($\texttt{cdfplot(data)}$) reaches level $q$, but not quite. The median is the $50^{th}$ percentile, the upper fourth is the $75^{th}$ percentile, the lower fourth is the $25^{th}$ percentile. It is useful to think of the $q^{th}$ sample quantile as the $x_(qn)$ data point in the ordered sample.

This procedure is not the only one in the literature or in current software.
Let us introduce the notation $\tilde{x}_q$ for the $q$th sample quantile of a random sample $x_1, \ldots, x_n$ (a random sample is a collection of independent random variables from the same distribution function $F$). This notation fits in with the notation $\tilde{x}$ for the sample median or $q = .50$ sample quantile.

Then $\tilde{x}_q \rightarrow F^{-1}(q)$ the $q$th quantile of the distribution, denoted $\eta(q)$ in our book as the size of the sample $n$ gets large. This is a consequence of the law of large numbers. For motivation, think of $\tilde{x}_q$ as approximately $F_n^{-1}(q)$, the point where the empirical cdf $F_n$ hits height $q$ (Recall: $F_n(x)$ is the fraction of data points less than or equal to $x$, and its graph jumps up by height $1/n$ at each data point.) Then $F_n \rightarrow F$ by the law of large numbers, and $\tilde{x}_q = F_n^{-1}(q) \rightarrow F^{-1}(q)$.

Then for a large sample $n$, one should see that $\tilde{x}_q \approx F^{-1}(q)$. The ordered data point $x_{(i)}$ is the $(i - .5)/n$ sample percentile, so the pairs

$$(F^{-1}(\frac{i - .5}{n}), x_{(i)})$$

should be close to the line $y = x$. A plot of these pairs is called a “probability” plot in our book, or sometimes a qqplot.

Recall that the quantiles $\eta(q)$ for a $N(\mu, \sigma^2)$ distribution are related to those of the $N(0,1)$ distribution with cdf $\Phi$ by

$$\eta(q) = \mu + \sigma \Phi^{-1}(q).$$

In other words, the graph $(\Phi^{-1}(q), \eta(q))$ is a straight line with slope $\sigma$ and intercept $\mu$.

If the underlying distribution behind the random sample is any $N(\mu, \sigma^2)$ distribution, then the graph using the $N(0,1)$ quantiles on the $x$-axis and the ordered data on the $y$-axis of the points

$$(\Phi^{-1}(\frac{i - .5}{n}), x_{(i)}), \ i = 1, \ldots, n$$

is useful without knowing $\mu, \sigma$. It is called the “normal probability plot” or “qqplot” (in Matlab the commands are qplot, normplot) . Since

$$(\Phi^{-1}(\frac{i - .5}{n}), x_{(i)}) \approx (\Phi^{-1}(\frac{i - .5}{n}), \mu + \sigma \Phi^{-1}(\frac{i - .5}{n}))$$

the plot will be near a straight line with slope $\sigma$ and intercept $\mu$ if the underlying distribution is Normal.