Principal Components Analysis, or in short PCA, is a way to take high dimensional data and reduce it without loosing much information. It is difficult to find clusters, similarities and differences in high dimensional data, but by doing PCA it becomes much easier because of the dimension reduction. [?]

If you have you multidimensional data plotted in a coordinate system, PCA finds the vectors which describes the most variance in the data. These are called the eigenvectors. The eigenvector describing the most variance is the first principle component, the eigenvector describing the second most variance is the second principle component and so on. The principle components are perpendicular. To find out which variables have the most influence on a PC, we can look at the coefficients of the PC, meaning the coefficients of the eigenvector, the higher the coefficient, the higher the influence. When having found the principle components, it is time to find out which to keep. This can bee done by looking at a scree plot. This is a bar chart showing the variance for each of the PC's. Typical it is interesting to look at the first two PC's. We then look at the data expressed in terms of the principle components we have chosen. If we plot the derived data as a scatter plot, with for example the first PC on the x-axis and the second PC on the y-axis, we can easily point out the clusters of the data if there are any. This way we can look at the data in fewer dimensions but without much loss of information.