**Principal component analysis**

What is PCA?  
Principal component analysis is an algorithm usually associated with machine learning in the data preprocessing phase, in which it used to reduce dimensionality of the data that are too large to smaller data that contain the most important information from the bigger data.  
Such action is performed to reduce the complexity of the model or increase its simplicity.

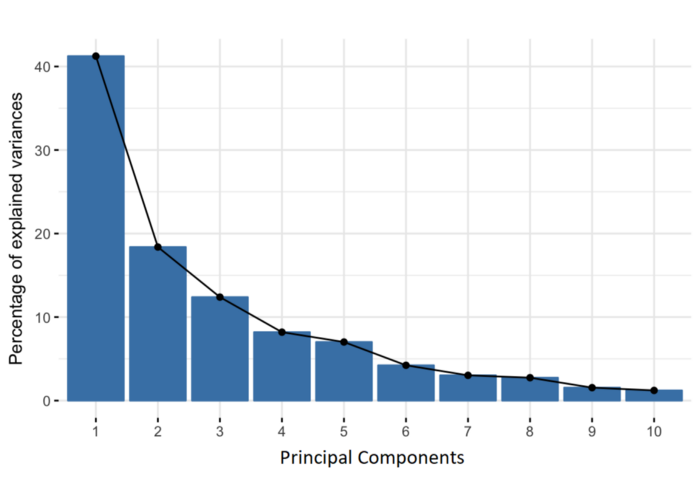
What exactly is the PCA output?  
PCA outputs new variables created by combining the original variables, these combinations are made so that the new variables are uncorrelated (not connected) and capture most of the information from the original variables.  
So if u have 10 dimensions, you get 10 PCA components arranged so that the first component contains the most important information, the second holds the second most important information and so it goes.  
Organizing the information this way will allow us to reduce the non-important information without losing much value.  
It's Also important to mention that the highest value of information points to the variance.

Fig.1 example on PCA components

Calculations of PCA:  
First, Standardization:  
The aim of this step is to standardize the data so that each of them contributes the same to the analysis, meaning we try to change the mean of the data to zero and the standard deviation to one.

Second, Covariance Matrix Calculation:  
This “Covariance Matrix” shows how correlated the components are, sometimes correlated data contribute more to the mean and the variance of the data in a way that they contain noises (redundancy).  
This matrix is a symmetric matrix that contains all the possible pairs of the initial data, it also shows us the relationship between the data, for example if the covariance between two initial variables is negative, it means that these variables are inversely related, if positive it means that they are directly related.

Third, Compute The Eigen Values And Eigen Variables of the Covariance Matrix:  
Eigen vectors represent the direction of the variance and the Eigen values are simply coefficient attached to them to represent the amount of variance represented in each component. By ordering the Eigen vectors according to the Eigen values we get the PCA components in order.