Principal component analysis:

* Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
* It is mostly used tool for making predictive models.
* It is used when there are too many variables and we need to find co-relation between these variables.
* Used to transform high dimensional data into low dimensional data.
* For example:
* (if we have n observations If we have dataset with only one variable (1-Dimentional graph) then we can plot it on number line and find the co-relation.
* If we have dataset with two variables (2-Dimentional graph) then we can plot it on x and y axis and find the co-relation.
* If we have dataset with three variables (3-Dimentional graph) then we can plot it on x, y and z axis and find the co-relation.
* But for 4-Dimension we can’t plot it on paper and for like 200 variables we need 200-Dimensional graph which is not possible.)
* Some dimensions are more important than others.
* Hence PCA takes data with lots of Dimensions and flattens it into 2-D or 3-D so that we can look at it.
* So how this algorithm works?
* For each variable we have to plot a PC.
* PC is a line which span the variation among the values.
* PC1 span most variation, PC2 span 2nd most variation and so on
* These PC1 and PC2 becomes our new axis
* So, till now we have plotted observations but we have to plot the variables
* We score genes on how much they influence PC
* Now we plot variable by using formula summation of (read count \* influence) for all observations
* Variables with similar pattern can be clustered togrther
* Influence-loading
* Array of loading-eigen vector