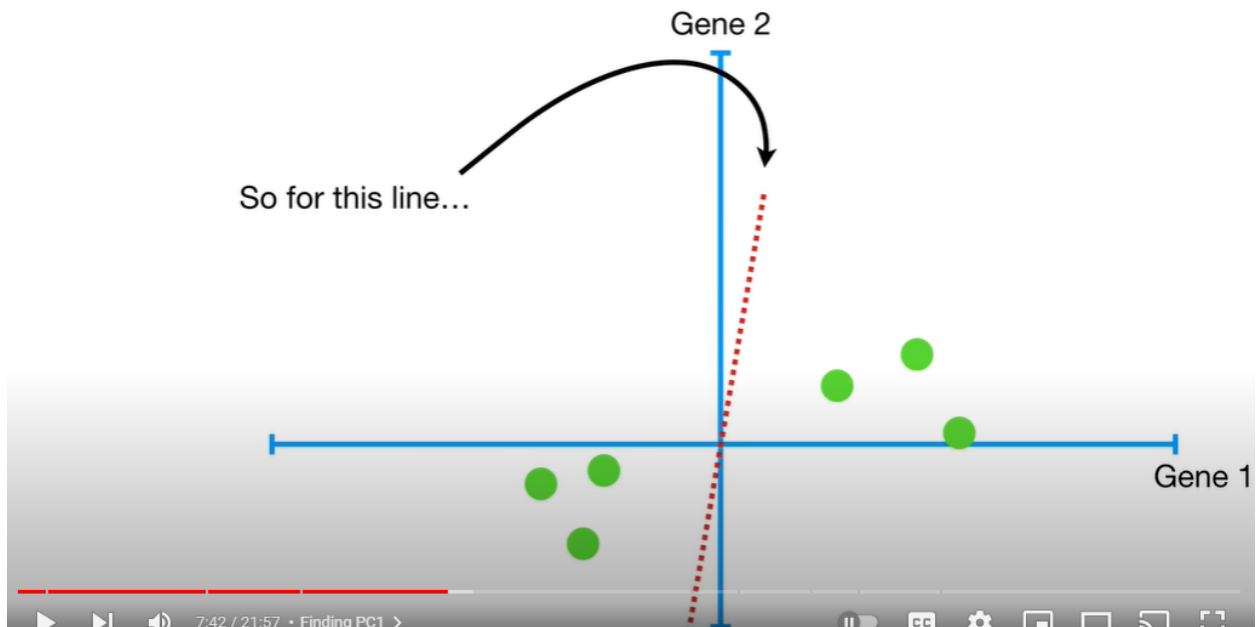


Catatan PCA

Video tersebut menjelaskan tentang bagaimana PCA dilakukan langkah demi langkah oleh StatQuest. PCA digunakan untuk mendapatkan wawasan yang lebih dalam tentang data dengan menggunakan SVD. PCA membantu mengelompokkan data ke dalam komponen utama yang menjelaskan variasi data dengan baik.

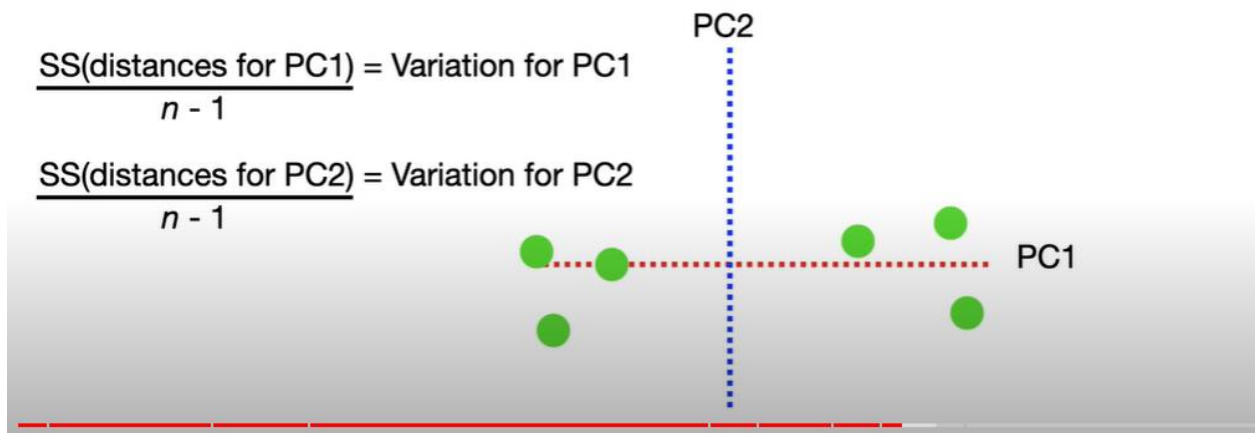


PCA menggunakan SVD untuk mengubah data ke dimensi yang lebih rendah, memungkinkan visualisasi dan pemilihan fitur yang lebih baik.

Well, if you are familiar with the equation for variation, you will notice **Eigenvalues** are just measures of variation.

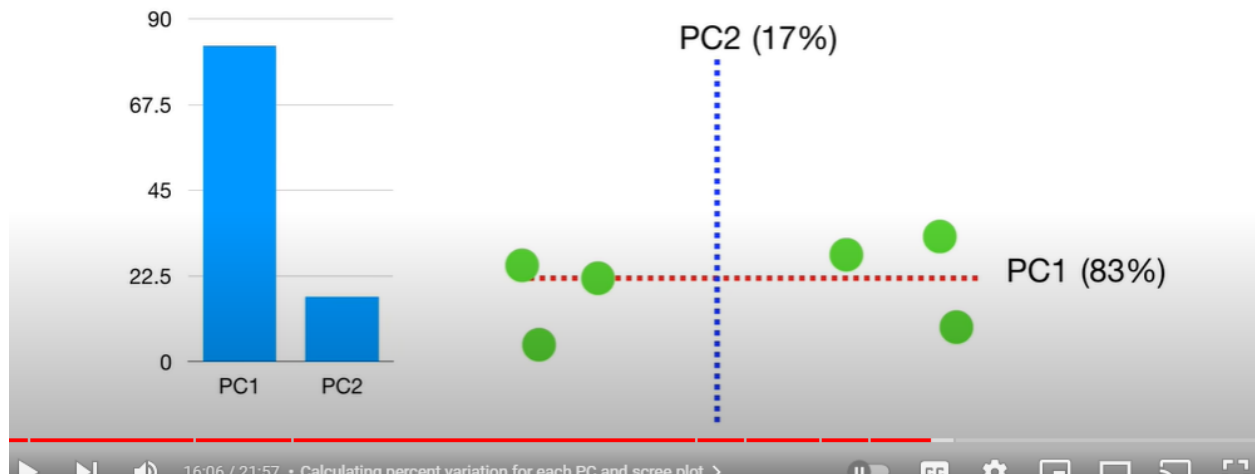
$$\frac{SS(\text{distances for PC1})}{n - 1} = \text{Variation for PC1}$$

$$\frac{SS(\text{distances for PC2})}{n - 1} = \text{Variation for PC2}$$



PCA menggunakan sum of squared distances untuk menemukan PC1 dan PC2, garis s yang merepresentasikan sebaran data PC1 memiliki sum of squared distances terbesar, sementara PC2 tegak lurus terhadap PC1.

TERMINOLOGY ALERT!!!! A Scree Plot is a graphical representation of the percentages of variation that each PC accounts for.



Scree plot adalah representasi grafis persentase variansi yang dijelaskan oleh setiap komponen utama. PC1 dan PC2 umumnya menyumbang mayoritas variasi.