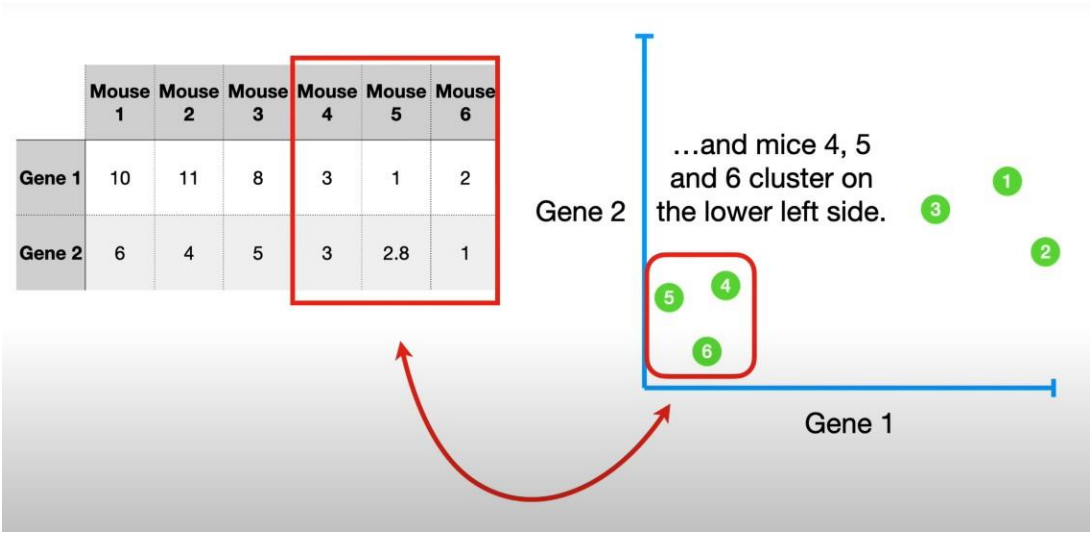


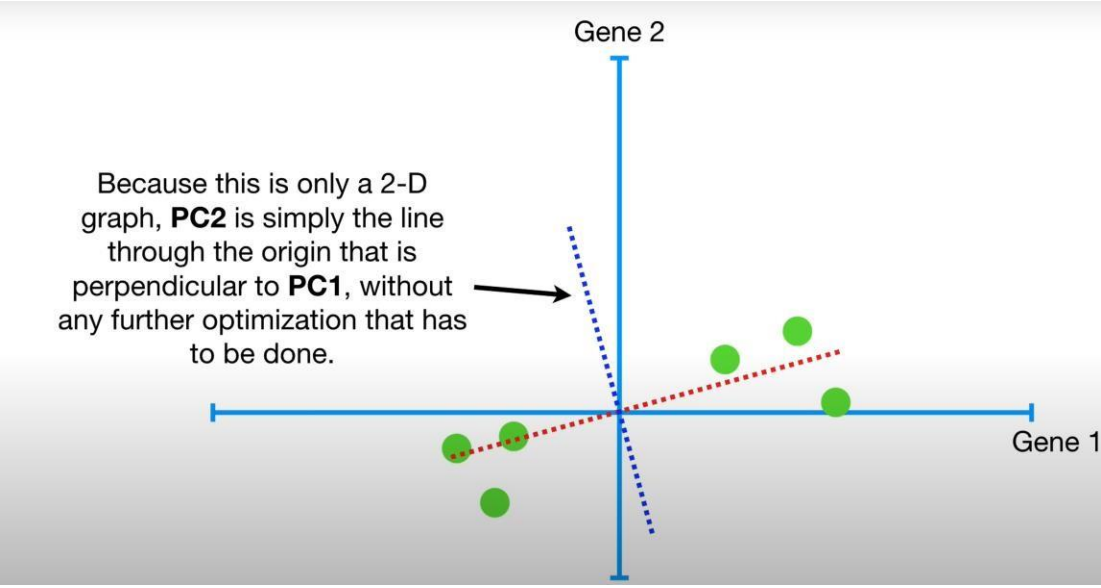
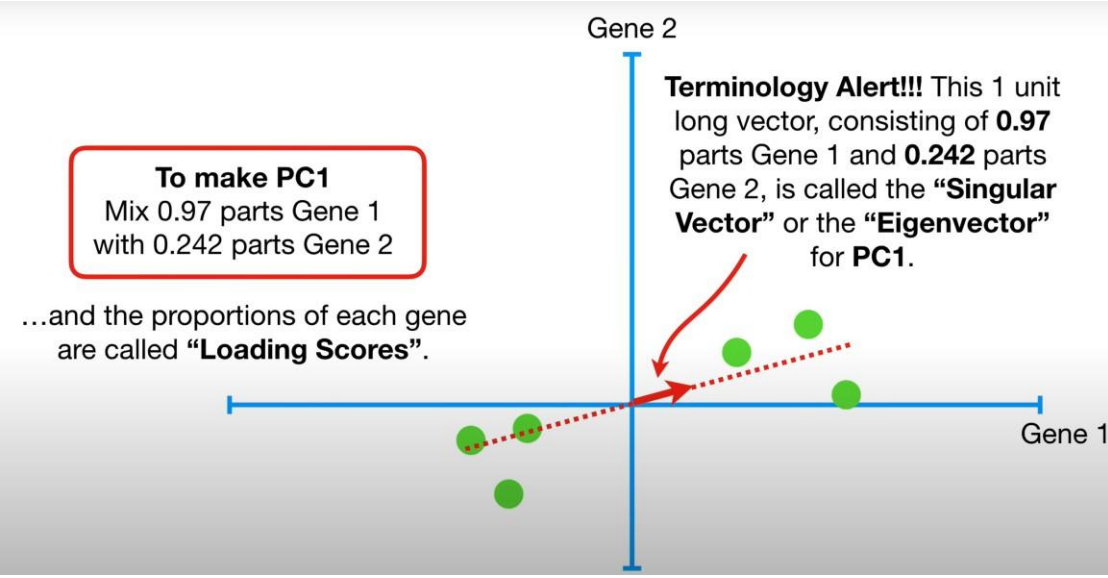
Rangkuman video  
StatQuest: Principal Component Analysis (PCA), Step-by-Step

Video tersebut menggunakan contoh data pengukuran transkripsi dua gen pada enam tikus yang berbeda. Dalam contoh ini, Gen 1 dan Gen 2 dijadikan sebagai variabel yang diukur pada setiap sampel. Proses pengukuran satu gen dapat direpresentasikan dengan menggunakan garis bilangan, sedangkan pengukuran dua gen direpresentasikan dalam grafik XY dua dimensi. Lebih lanjut, pengukuran tiga gen menghasilkan grafik 3D, sementara pengukuran empat gen membutuhkan dimensi yang lebih tinggi, yakni 4 dimensi, yang akan dibahas melalui PCA.



PCA berfungsi untuk membantu mengelompokkan data dan mengidentifikasi variabel yang paling berharga dengan cara memproyeksikan data ke garis dan mengukur jaraknya untuk menemukan garis yang meminimalkan jarak dari titik proyeksi ke titik asal. Proses perhitungan PCA melibatkan penghitungan pusat data dan memproyeksikannya ke garis acak, di mana garis yang paling sesuai ditemukan dengan meminimalkan jumlah kuadrat jarak dari titik proyeksi ke titik asal.

Komponen Utama (PC) terdiri dari PC1, yang adalah garis yang paling sesuai dengan data, dengan kemiringan yang menunjukkan proporsi pentingnya masing-masing variabel, serta PC2, yang adalah garis tegak lurus dengan PC1. Nilai eigen digunakan sebagai ukuran variasi yang dijelaskan oleh setiap PC, di mana PC1 dan PC2 menggambarkan sebagian besar variasi data. Grafik PCA digunakan untuk menggambarkan sampel-sampel dalam dimensi yang lebih rendah, dan jika hanya ada dua PC utama, grafik PCA dua dimensi cukup informatif.



Kesimpulannya, PCA membantu dalam mereduksi dimensi data dan mengidentifikasi pola yang penting, sementara scree plot digunakan untuk menentukan jumlah PC yang signifikan. Terlebih lagi, PCA tetap dapat digunakan bahkan jika tidak memungkinkan untuk menggambar grafik dalam dimensi yang sesuai.