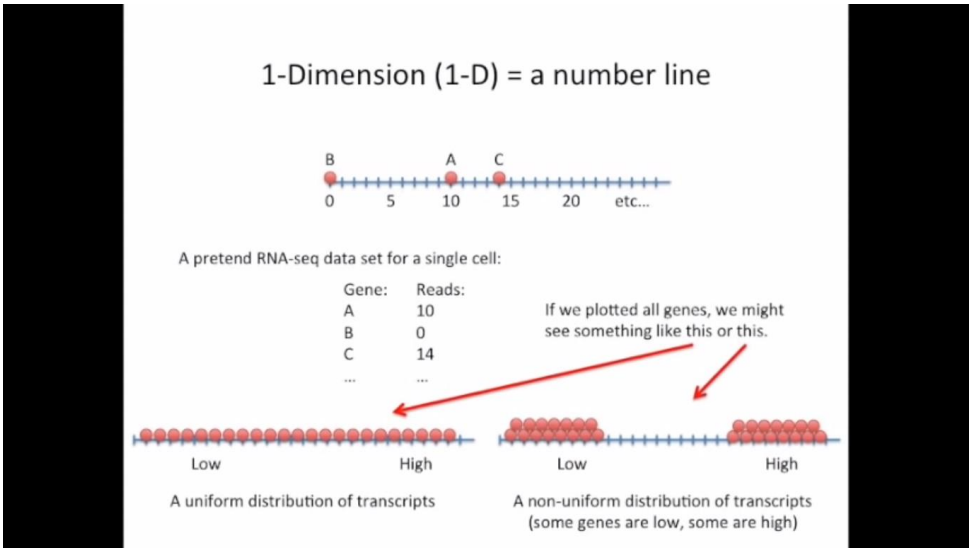
Nama : Ardhien Fadhillah Suhartono

NIM : 1103204137

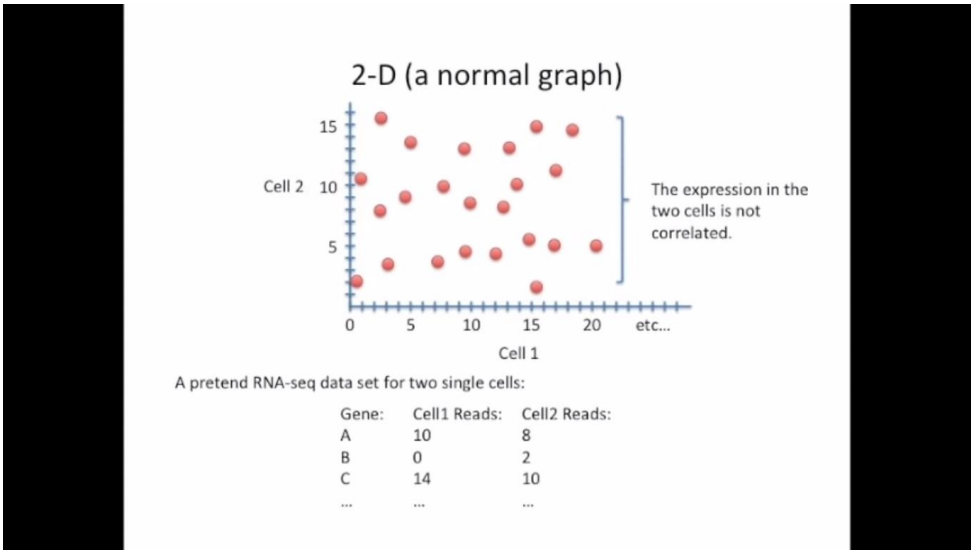
Kelas : TK4402

Understanding 3 link statquest

1. PCA 1 Dimension
   Merupakan data variabel yang paling sederhana, dimana hanya akan diwakiliki oleh satu variabel dalam satu garis linear yang dapat digunakan sebagai pengurang dimensi data tanpa kehilangan terlalu banyak informasi.

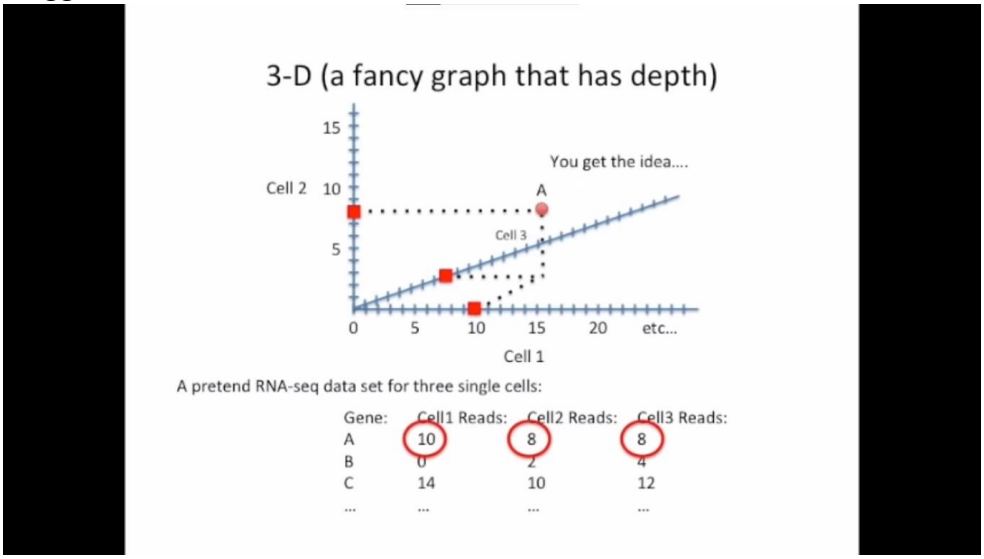


2. PCA 2 Dimension
   Sama seperti 1-D akan tetapi memilki 2 garis cell yang dimana bisa diwakilkan hingga 2 variabel.



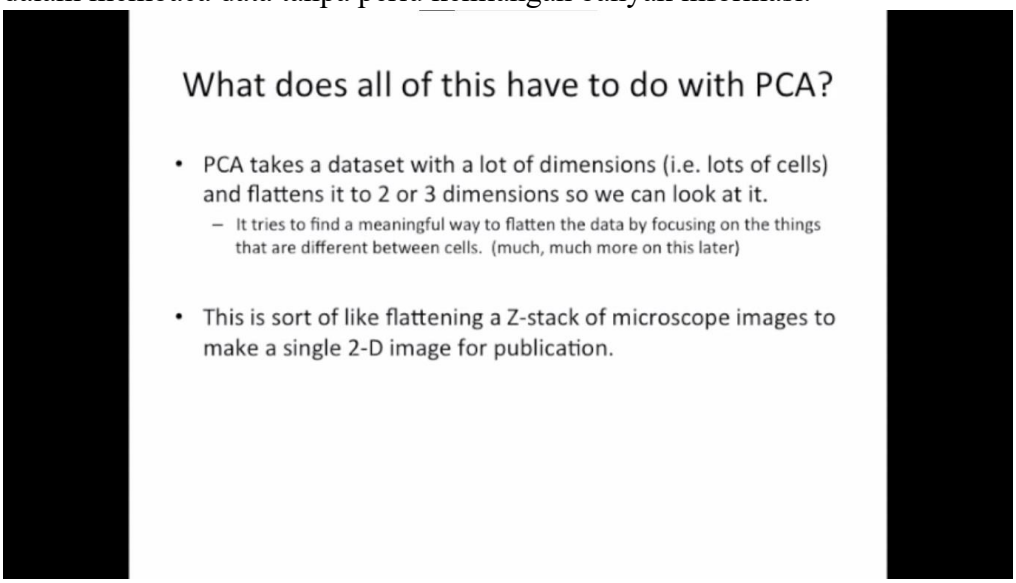
3. PCA 3 Dimension

Sama seperti sebulumnya akan tetapi lebih komplekks dikarenakan damapat mewakili hingga 3 variabel berbeda

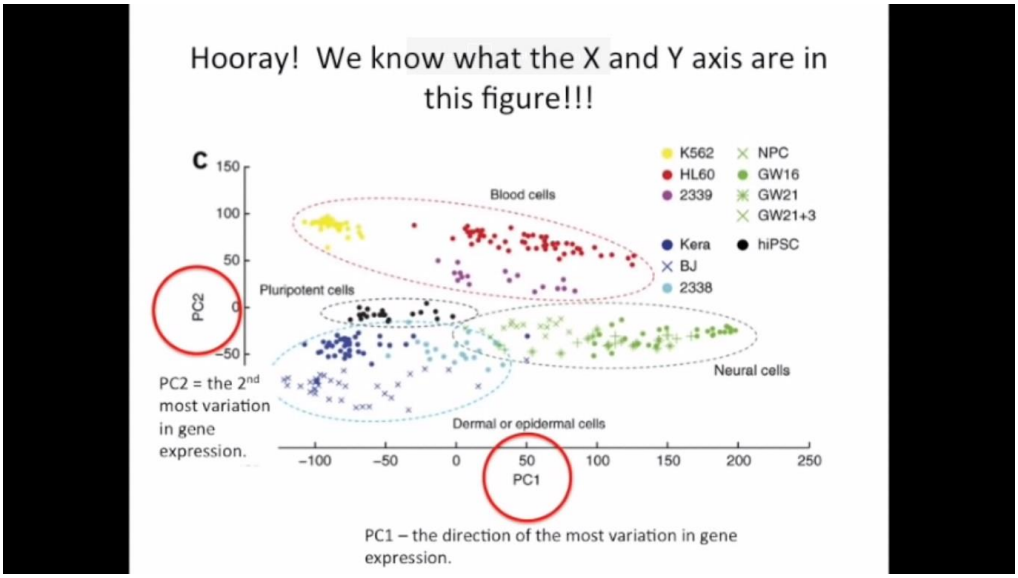


Why We Can Omit Dimensions

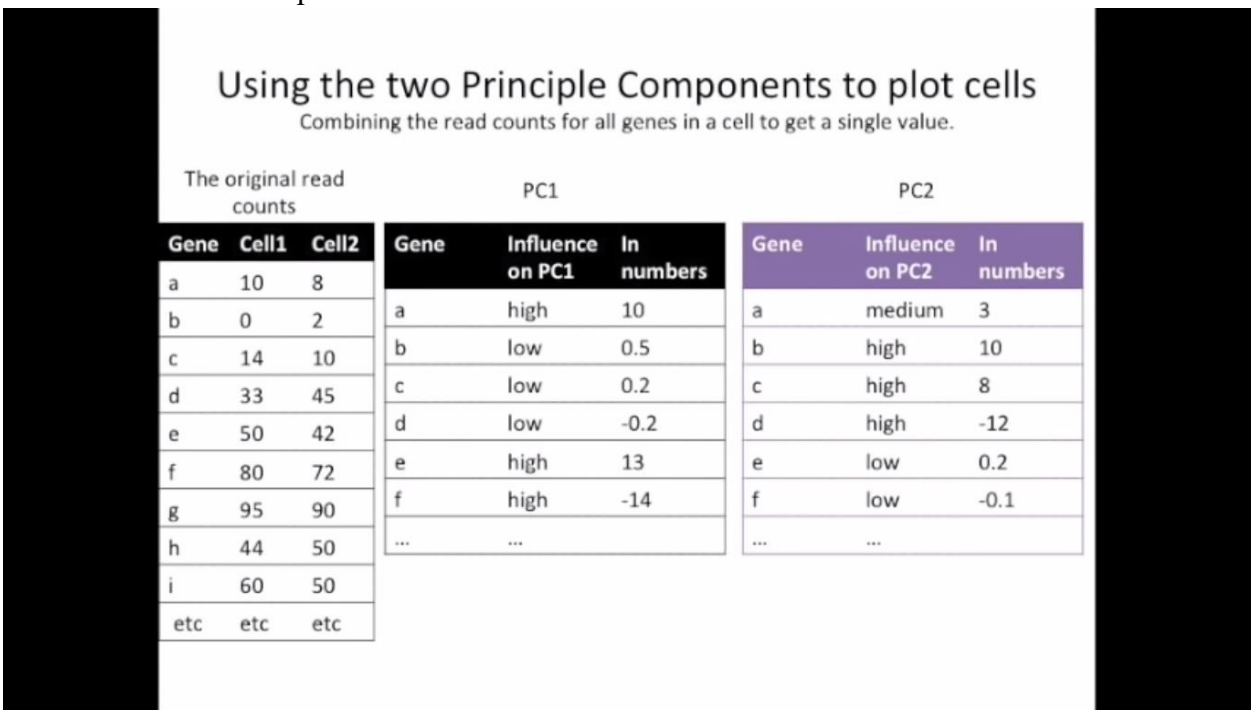
1. Dijelaskan bahwa kita dapat mengedit data sesuka kita, dimana ini akan memudahkan dalam membaca data tanpa perlu kehilangan banyak informasi.



2. Pengelompokan berdasarkan dimensi memberikan representasi utama dari variasi data. Misalnya, PC1 mencerminkan variasi terbesar dari kiri ke kanan, sementara PC2 mencerminkan variasi kedua dari atas ke bawah. Untuk memahami pengaruh gen terhadap penempatan sel, perhatikan skor pengaruhnya di PC1; sedangkan untuk membedakan sel darah, sel saraf, dan sel kulit, tinjau skor pengaruh di PC2.



3. Untuk membuat plot sel menggunakan komponen utama, tentukan seberapa besar pengaruh suatu gen terhadap komponen utama, seperti PC1. Pada contoh PC1, gen yang paling berpengaruh berada di ujung kiri atau kanan, dikenal sebagai "extreme genes". Setelah menetapkan tingkat pengaruh gen, berikan nilai numerik dan hitung seluruh skor utama untuk membuat plot sel.



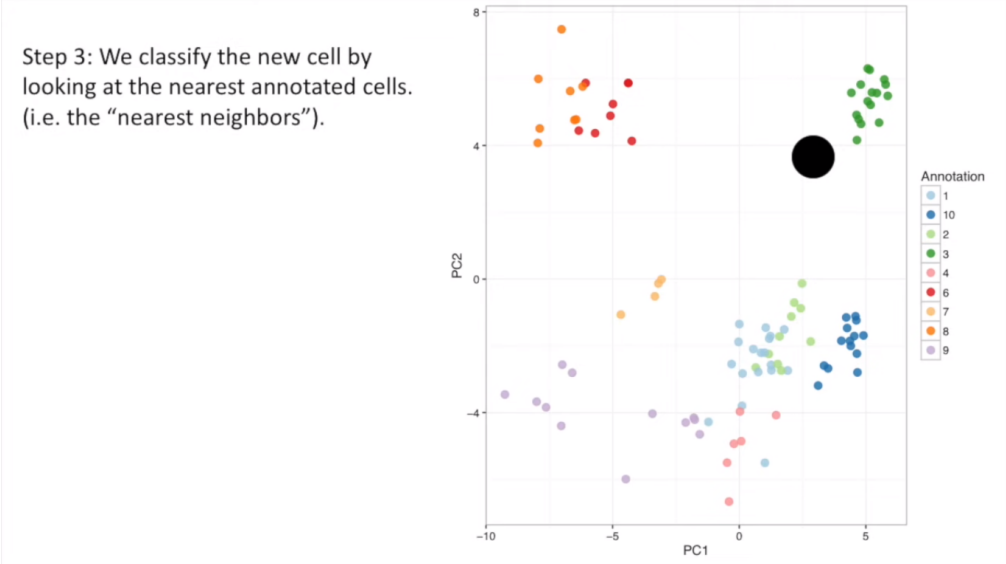
Gene	Cell1	Cell2
a	10	8
b	0	2
c	14	10
d	33	45
e	50	42
f	80	72
g	95	90
h	44	50
i	60	50
etc	etc	etc

Gene	Influence on PC1	In numbers
a	high	10
b	low	0.5
c	low	0.2
d	low	-0.2
e	high	13
f	high	-14
...	...	

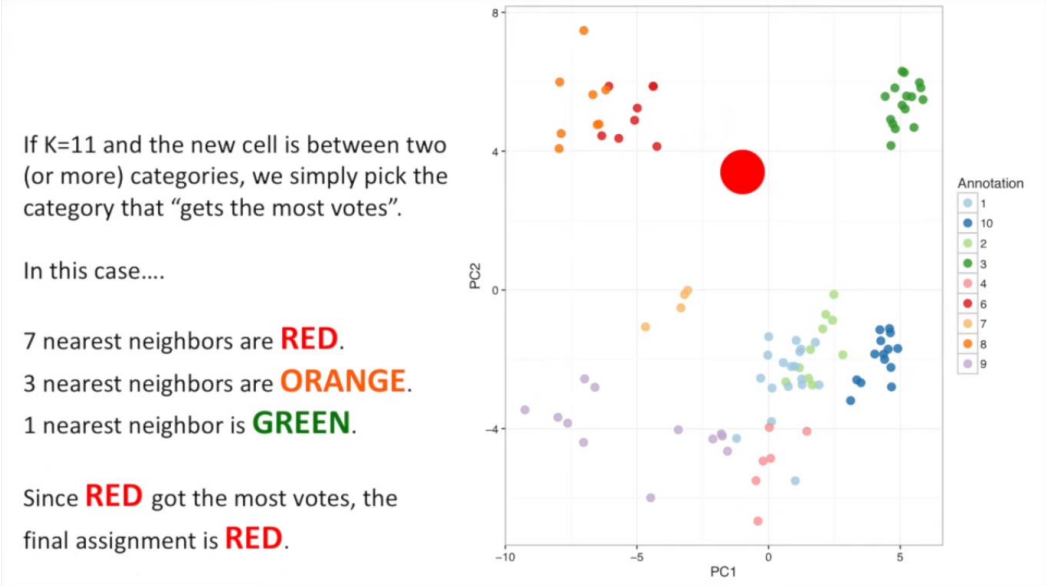
Gene	Influence on PC2	In numbers
a	medium	3
b	high	10
c	high	8
d	high	-12
e	low	0.2
f	low	-0.1
...	...	

KNN

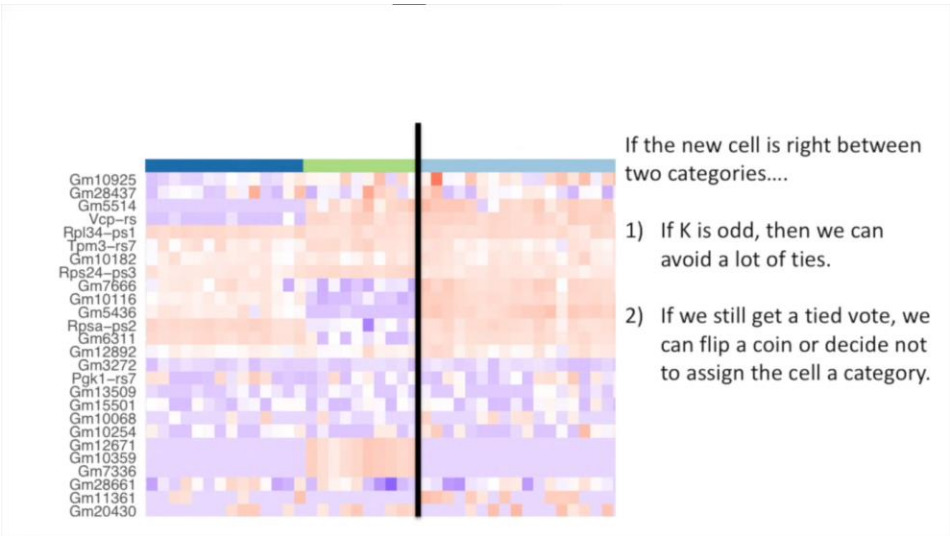
1. Disini diberikan contoh dari perkumpulan sel tumor yang sudah di klasifikasikan, lalu akan ditambahkan sel baru untuk dilihat bagaimana pergerakan sel tersebut



2. Disini dijelaskan bahwa sel asing tersebut akan mengambil warna sesuai dengan warna apa dia terdekat atau berapa banyak vote yang ia dapatkan dari warna sel sel lain

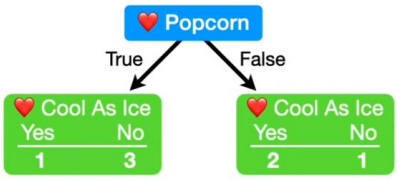


3. Sama halnya yang diatas pada contoh ini garis akan mengikuti dari warna apa di terdekat atau vote terbanyak, namun apabila terdapat di Tengah Tengah bisa dilakukan cara melempar koin



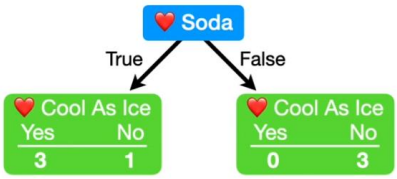
Decision and Classification Trees, Clearly Explained!!!

1. Percobaan ini mencatat preferensi terhadap popcorn, baik dalam versi "cool as ice" maupun tidak. Hal serupa juga dilakukan untuk orang-orang yang menyukai soda.

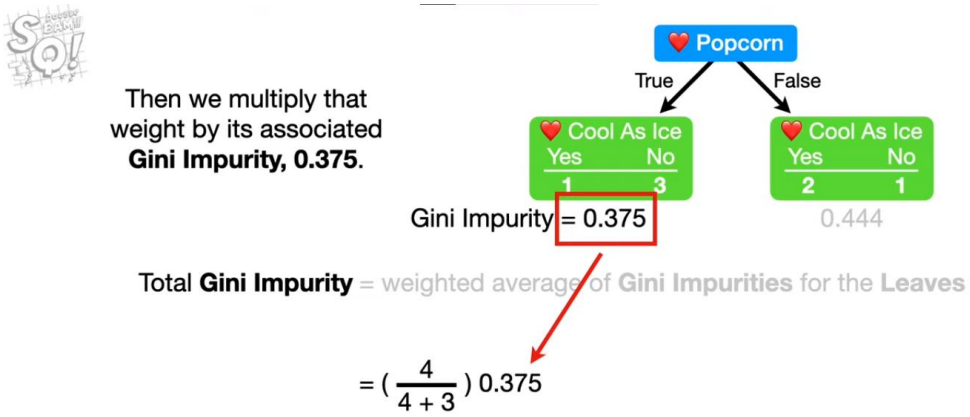


However, numerically, the methods are all quite similar*, so we will focus on **Gini Impurity** since, not only is it very popular, I think it is the most straightforward.

* For details, see page 321 of the Introduction to Statistical Learning in R.



2. Dalam algoritma decision tree, Metode Gini Impurity berfungsi untuk mengukur sejauh mana ketidakmurnian suatu node. Jika impurity lebih tinggi, itu menunjukkan tingkat variasi yang lebih besar dan klasifikasi yang lebih sulit dilakukan. Oleh karena itu, Loves Soda ditempatkan di puncak pohon keputusan karena memiliki Gini Impurity terendah, menunjukkan kejelasan klasifikasi yang lebih baik.



3. Untuk membuat decision tree, kita harus menentukan pernyataan pertama dengan memilih yang memiliki Gini Impurity terendah. Impurity menunjukkan sejauh mana konsistensi jawaban dalam suatu pernyataan. Gini Impurity juga digunakan untuk menentukan cabang-cabang berikutnya, mirip dengan langkah pertama. Proses ini diulang hingga mencapai prediksi yang diinginkan. Misalnya, dalam kasus di mana "loves soda" memiliki impurity terendah pada bagian "false", menyiratkan bahwa data "does not love cool as ice" adalah kesimpulan di leaf tersebut. Begitu juga pada leaf dengan umur <12.5 tahun yang benar. Namun, pada kasus umur <12.5 tahun yang salah, kesimpulan diambil berdasarkan mayoritas hasil voting dari data di leaf, yaitu "Loves cool as ice".