

1. Principal Component Analysis

This PCA plot shows clusters of cell types.

This graph was drawn from single-cell RNA-seq.

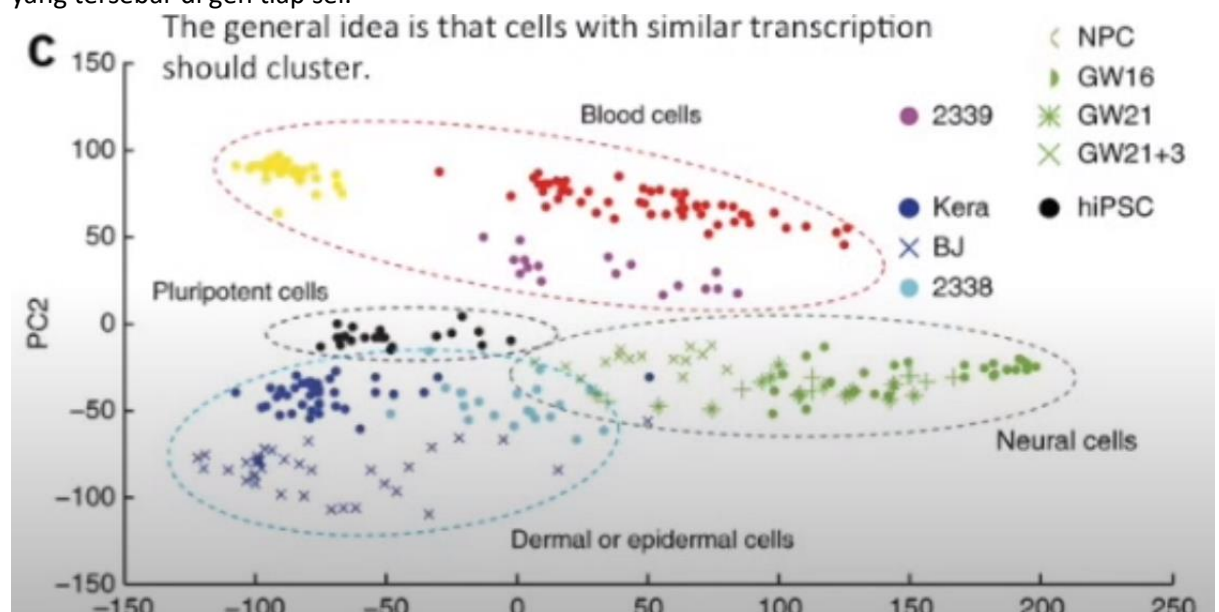
There were about 10,000 transcribed genes in each cell.

Each dot represents a single-cell and its transcription profile

The general idea is that cells with similar transcription

< NPC

Pada grafik tersebut menggambarkan mengenai bagaimana penggambaran single-cell RNA yang tersebar di gen tiap sel.



Titik menggambarkan single-cell

- 1 cell = 1-D graph (number line)
- 2 cells = 2-D graph (normal x/y graph)
- 3 cells = 3-D graph (fancy graph with depth)
- 4 cells = 4-D graph (you can't draw it)
- 200 cells = 200-D graph (etc..)

Menggambarkan bagaimana bentuk bentuk dari tiap dimensi grafik

General ideas so far...

- Genes with the largest variation between cells will have the most influence on the principal components.
 - i.e. genes highly expressed in some cells and not expressed in others will have a lot of variation and influence on the PCs.

Gen yang memiliki variasi yang besar antar sel akan memiliki most influence dari prinsip komponen.

2. StatQuest

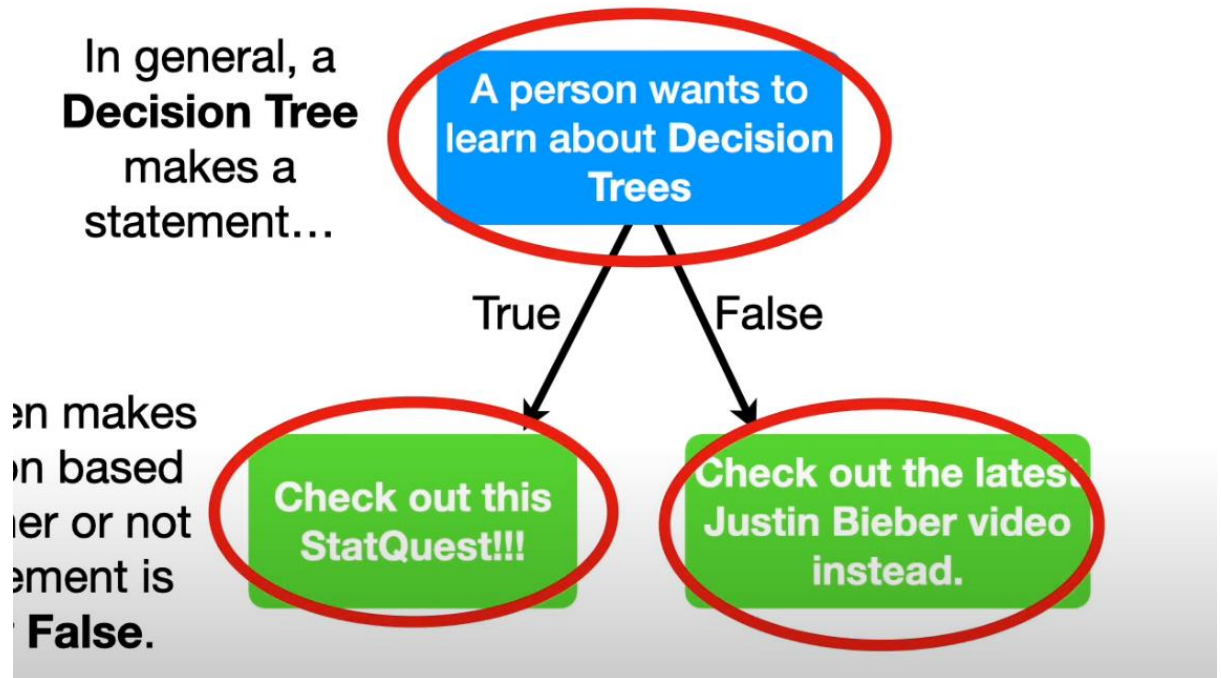
The K-Nearest Neighbors Algorithm

- A super simple way classify data.

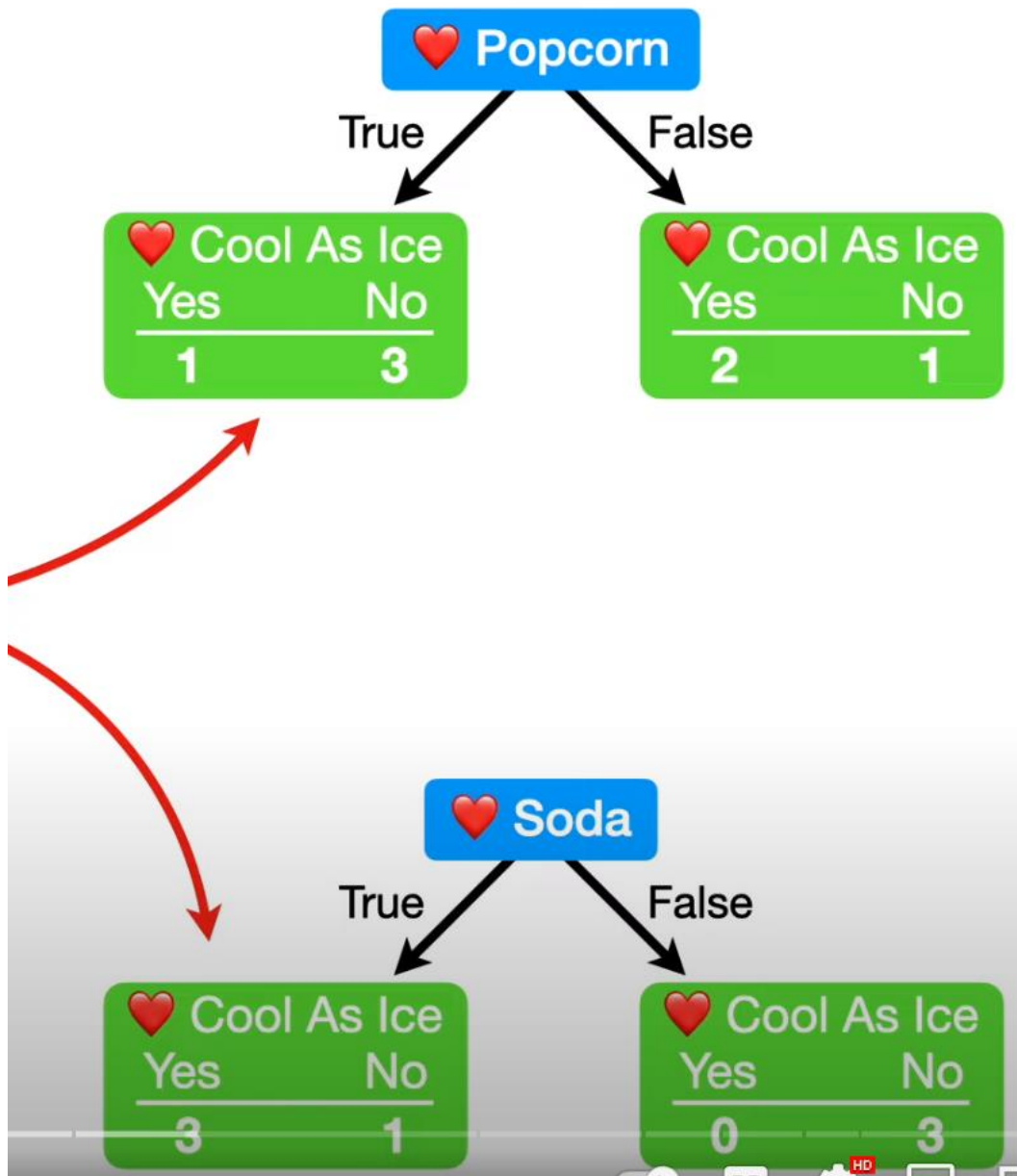
If you already had a lot of data that defined these cell types...

KNN ini adalah suatu algoritma yang memudahkan untuk mengclassify sebuah data. Intinya apabila ada suatu dataset yang unknown maka akan terjadi 2 hal dataset unknown tersebut akan mengikuti dataset terdekatnya atau mengikuti dataset yang terdapat banyak vote yang bertetangga.

3. Decision and Classification trees



Dalam sebuah tree decision akan ada sebuah pilihan bisa pilihan salah atau benar. Apabila a Decision Tree klasifikasi benda tertentu menjadi kategori maka disebut Classification Tree, namun apabila Decision treenya berupa numeric Values maka akan menjadi Reggresion Tree.



Disini Tree Decision menggunakan Gini Impurity dimana, algoritma ini digunakan untuk mengukur sejauh mana kemurnian node pada pohon.