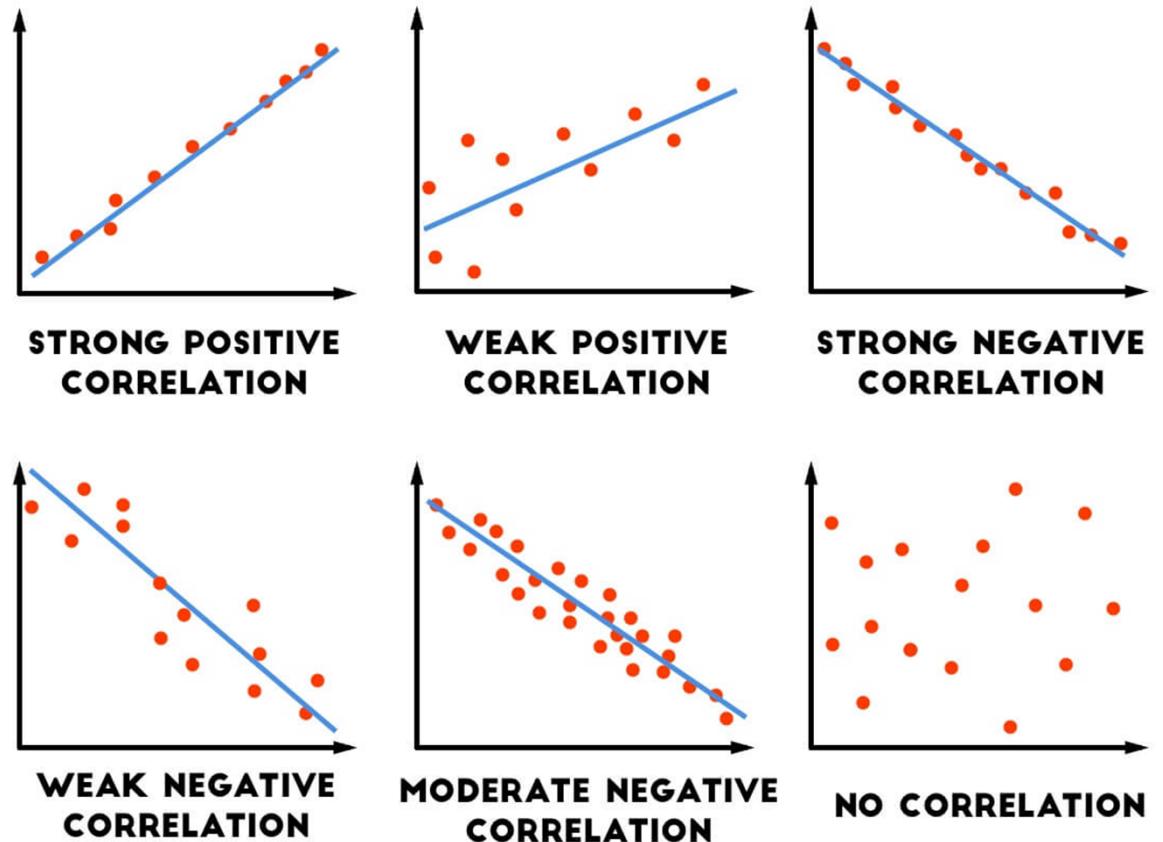


Correlation

It exists between two variables when the **values of one variable** are somehow **associated** with the **values of the other variable**.

It can be **positive, negative, non-existent or non – linear**.

Types of Correlation





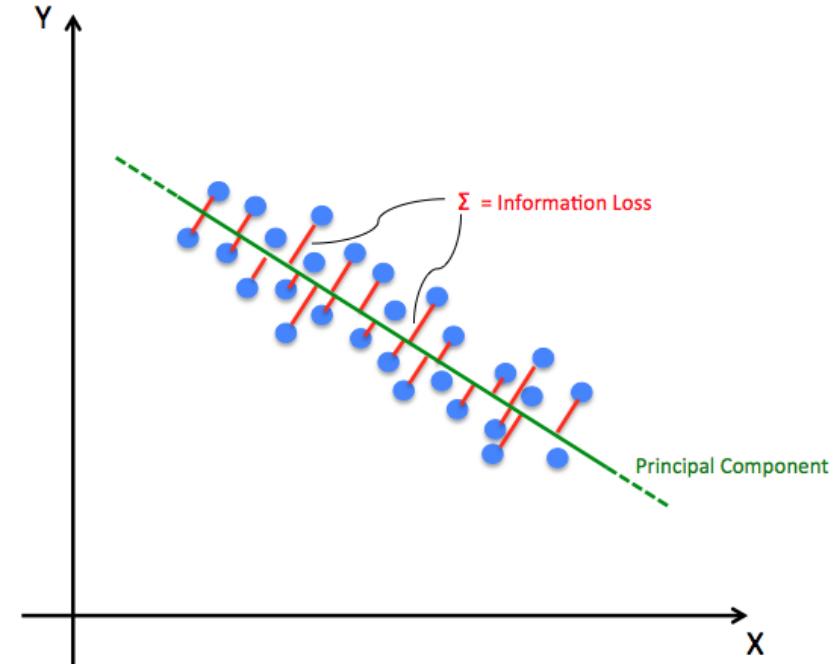
The movies are **correlated**



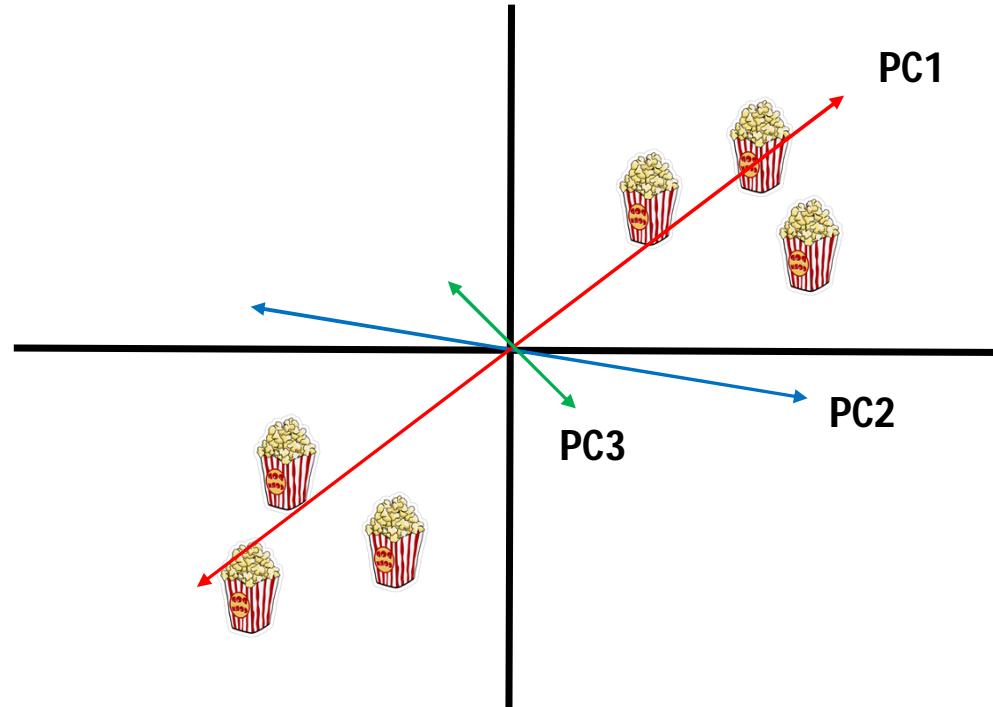
Also , Why is the line plotted diagonally?



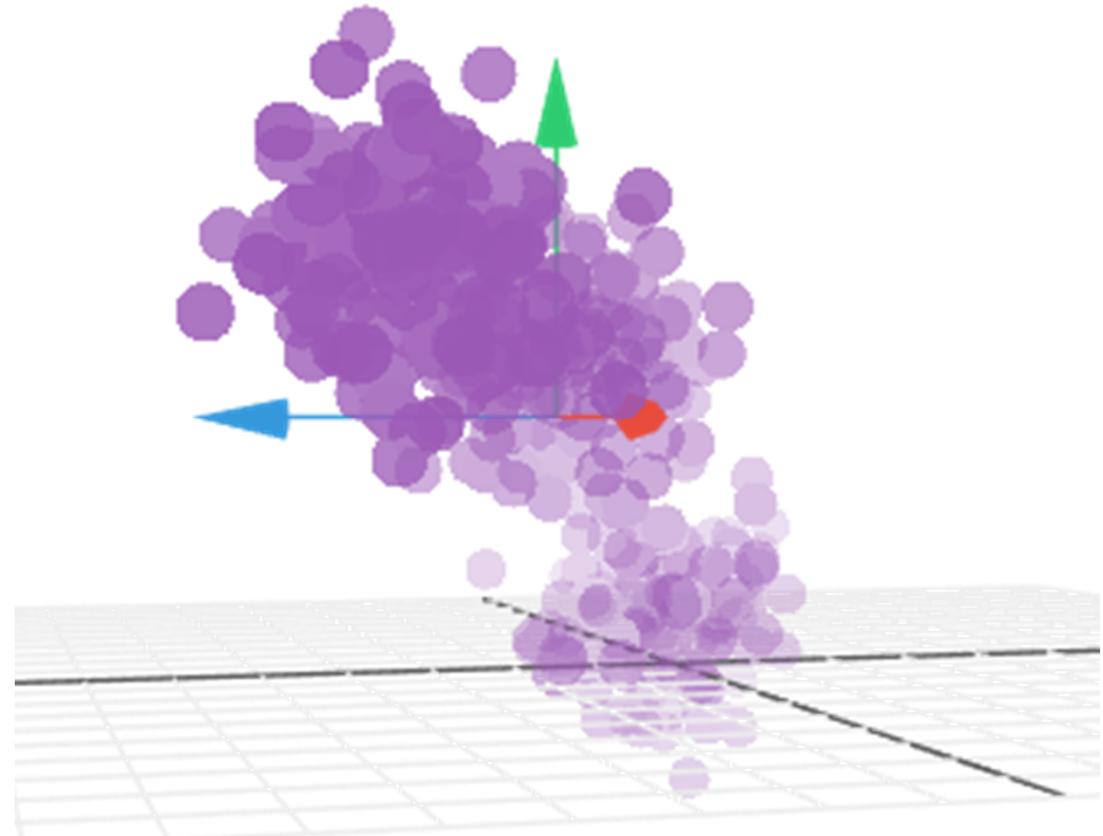
- The line is plotted in the direction where the **features are spread out the most or largest variance.**
- This is because it can retain the **maximum amount of information** in the original data.
- It is called **Principal component axis.**
- It **minimizes the information loss.**



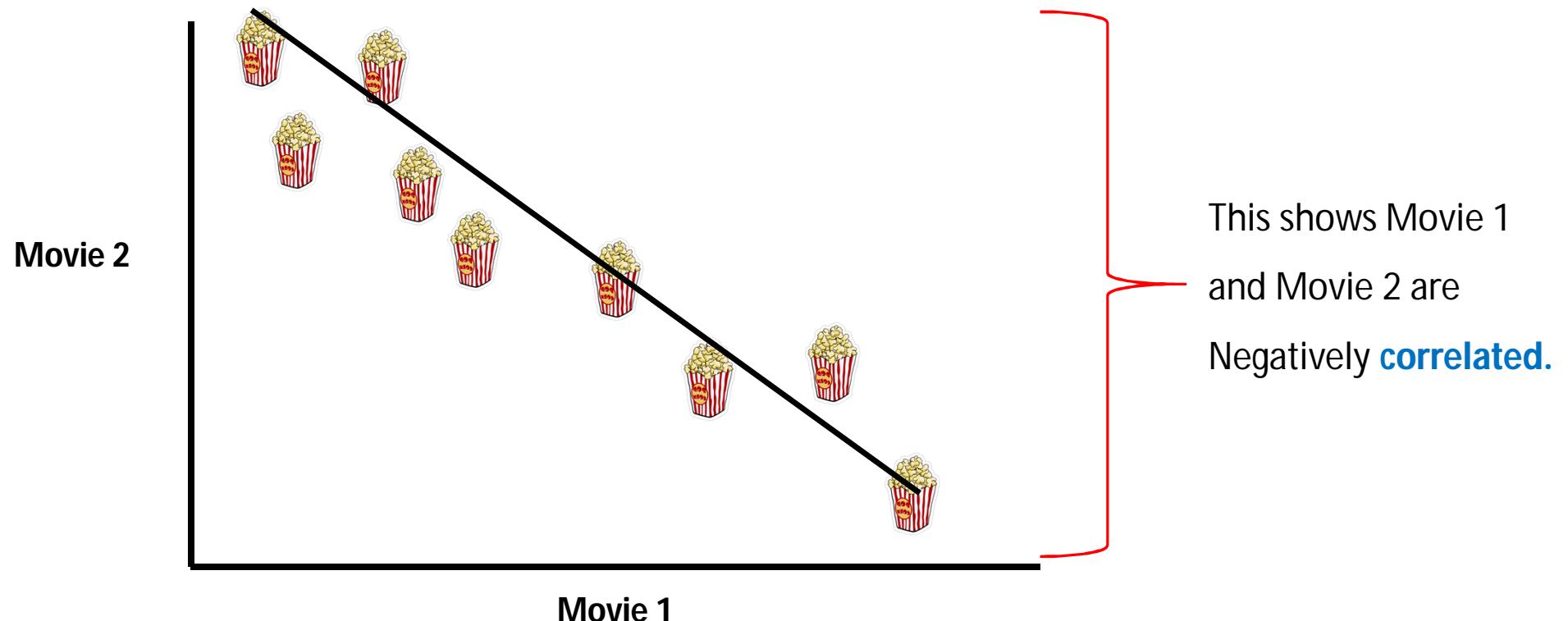
- Since PC1 captures the direction where **most of the variation** is.
- Therefore, it is the **best fitting line**.



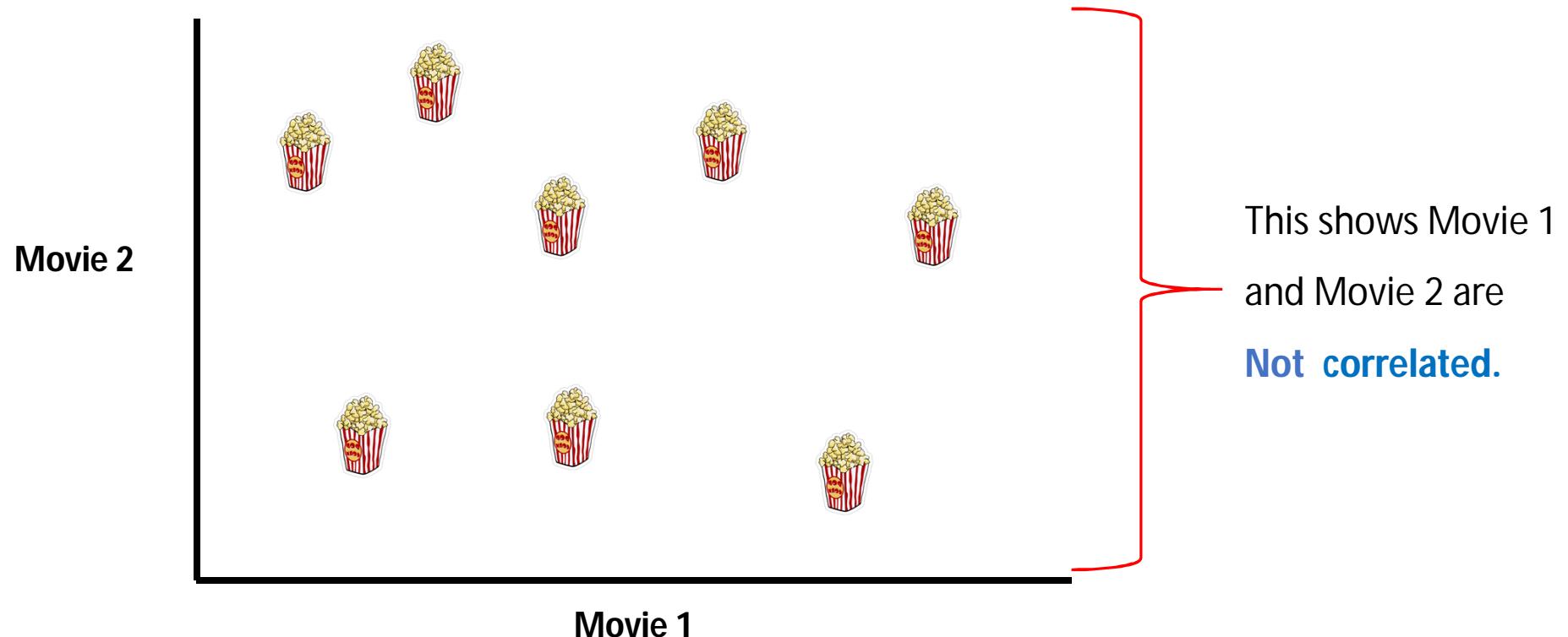
Graphical Visualisation



What if our graph looked like this ?



What if our graph looked like this ?



Lets consider three columns(Movie1, Movie2 & Movie3):

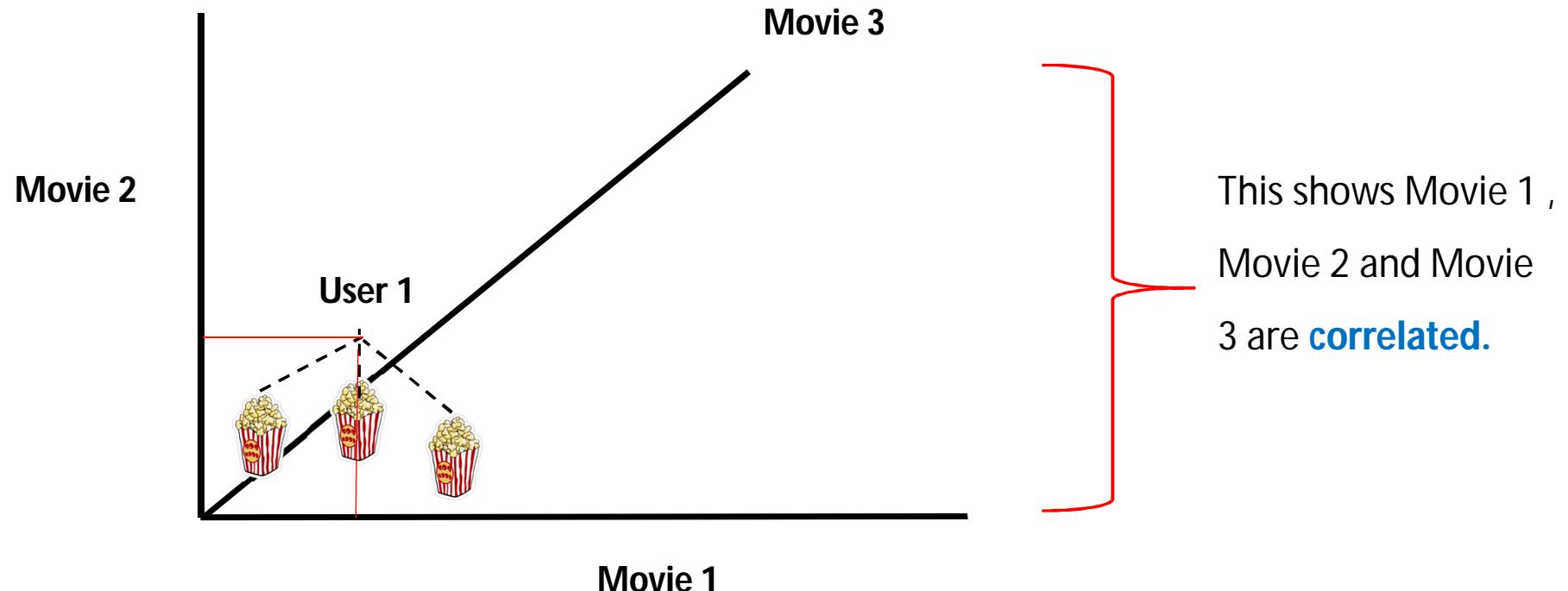
Movies (features) →

↓
Users (instances)

	Movie 1	Movie 2	Movie 3	Movie 4
Show 1	100	98	80	70
Show 2	0	20	40	60
Show 3	140	100	120	80
Show 4	330	450	480	520

After plotting our graph would look like:

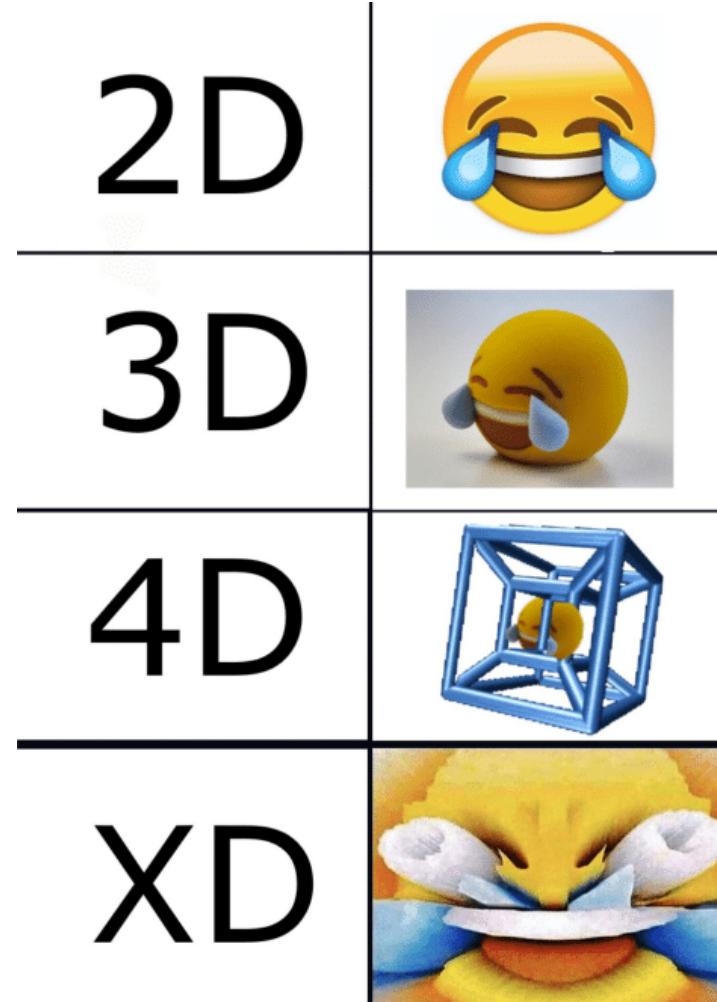
3 - D(Graph with Depth)





Instead we draw a **PCA plot** which converts the correlations among all of the movies to a **lesser dimension.**

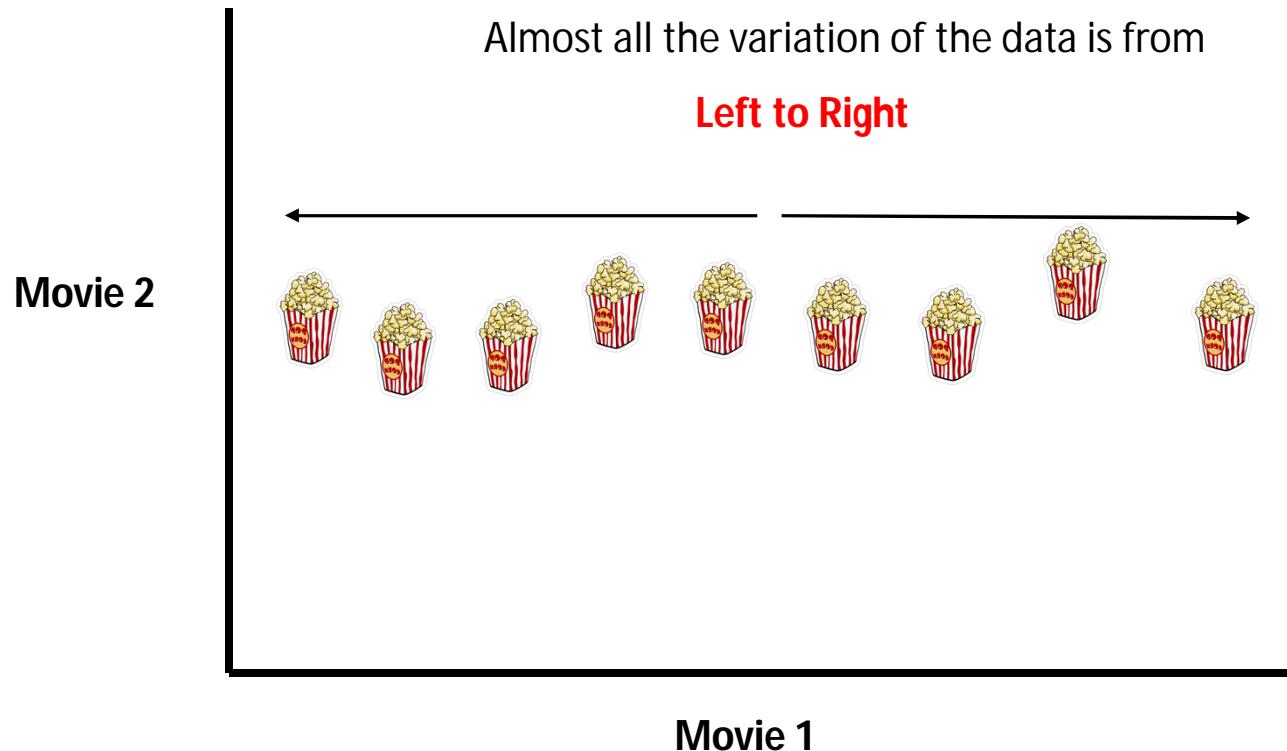




Lets get back to 2 D



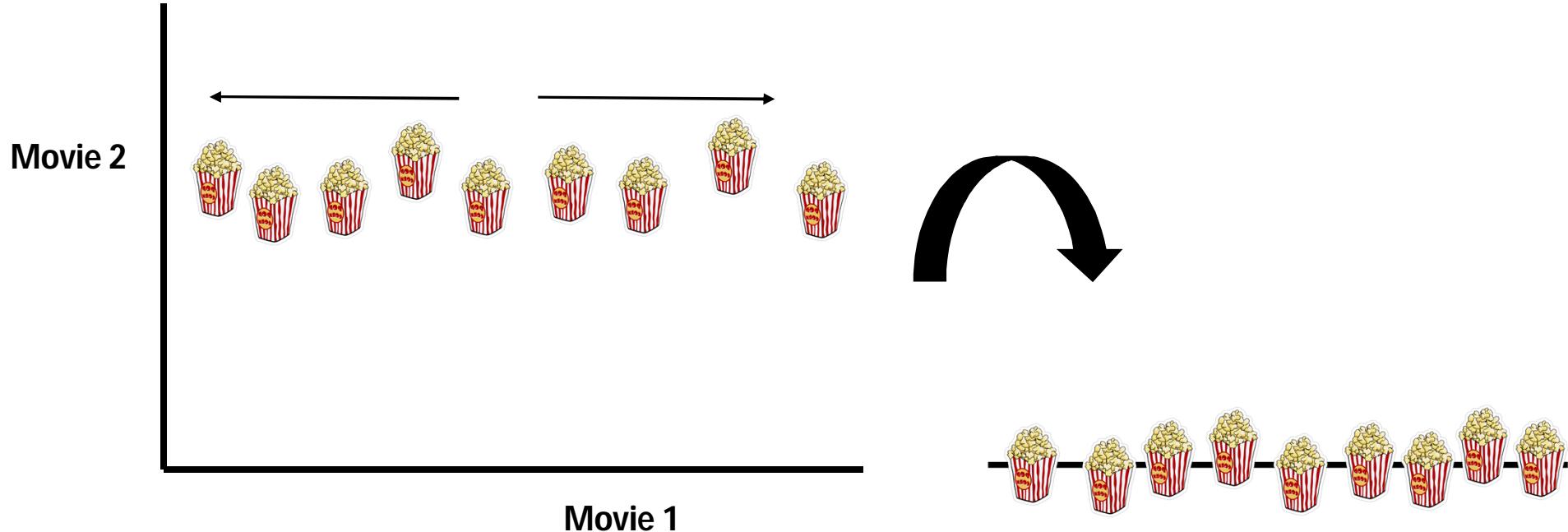
If our graph looked liked this :



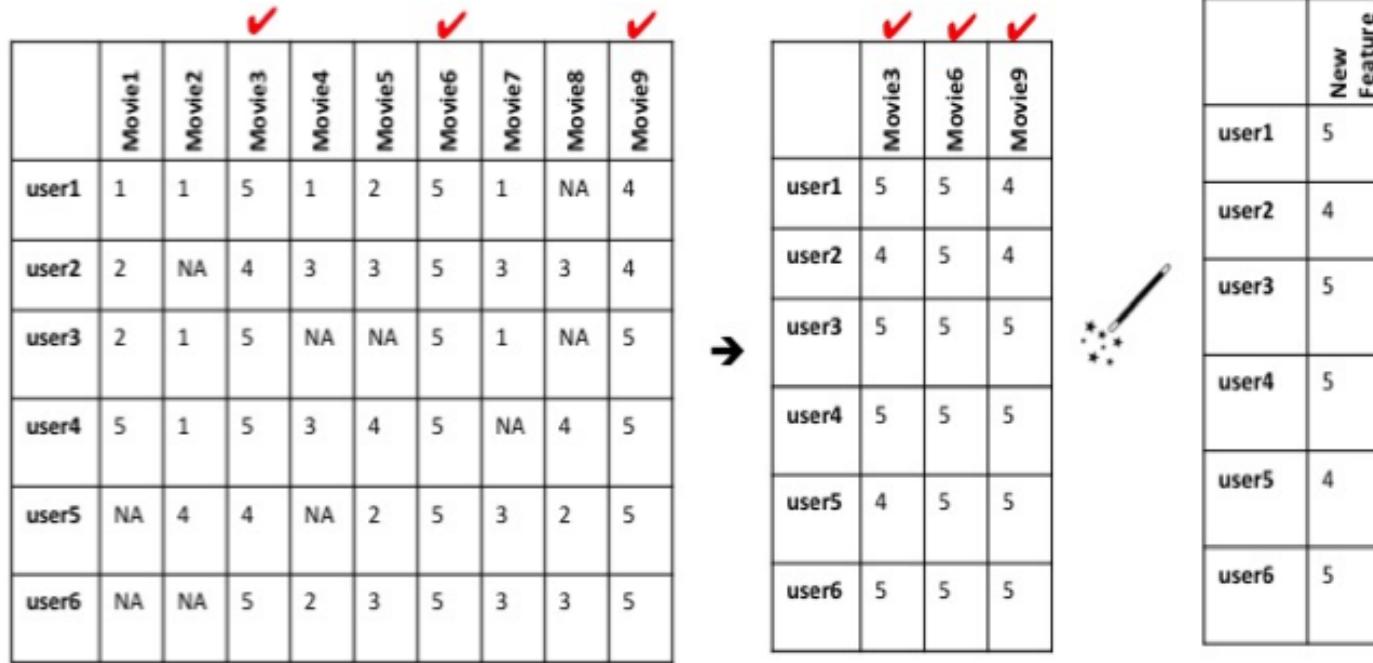
We can take 2 D data and display it on 1 D graph:

In both cases, the variation of the data is from

Left to Right



We can reduce data from **3D to 2D** and **2D to 1D** without much **loss of information**.



The diagram illustrates a process of data reduction. It starts with a 3D matrix of user ratings for nine movies. A red arrow points to a 2D matrix where rows are grouped by movie. Another red arrow points to a 1D vector where each row is represented by a single value.

	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	Movie7	Movie8	Movie9
user1	1	1	5	1	2	5	1	NA	4
user2	2	NA	4	3	3	5	3	3	4
user3	2	1	5	NA	NA	5	1	NA	5
user4	5	1	5	3	4	5	NA	4	5
user5	NA	4	4	NA	2	5	3	2	5
user6	NA	NA	5	2	3	5	3	3	5

	Movie3	Movie6	Movie9
user1	5	5	4
user2	4	5	4
user3	5	5	5
user4	5	5	5
user5	4	5	5
user6	5	5	5

	New Feature
user1	5
user2	4
user3	5
user4	5
user5	4
user6	5

Applications



For Images:

- It can be used for **image compression**.
- We can actually **reduce the size** of the image.