

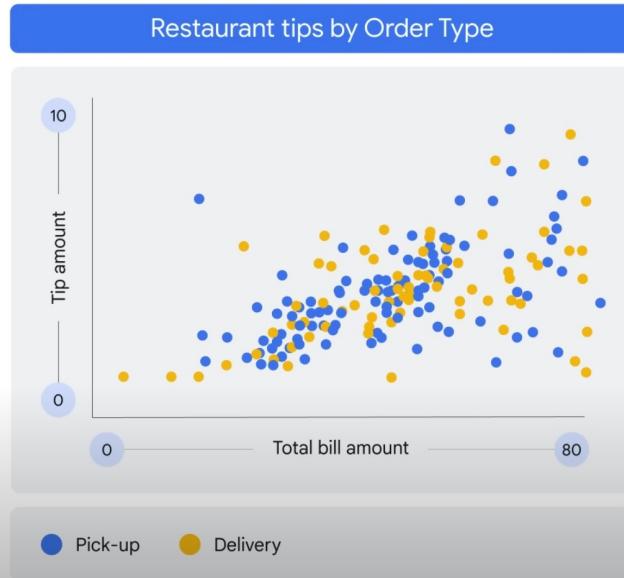
# THE BASICS OF UNSUPERVISED LEARNING

SOUMYA BANERJEE

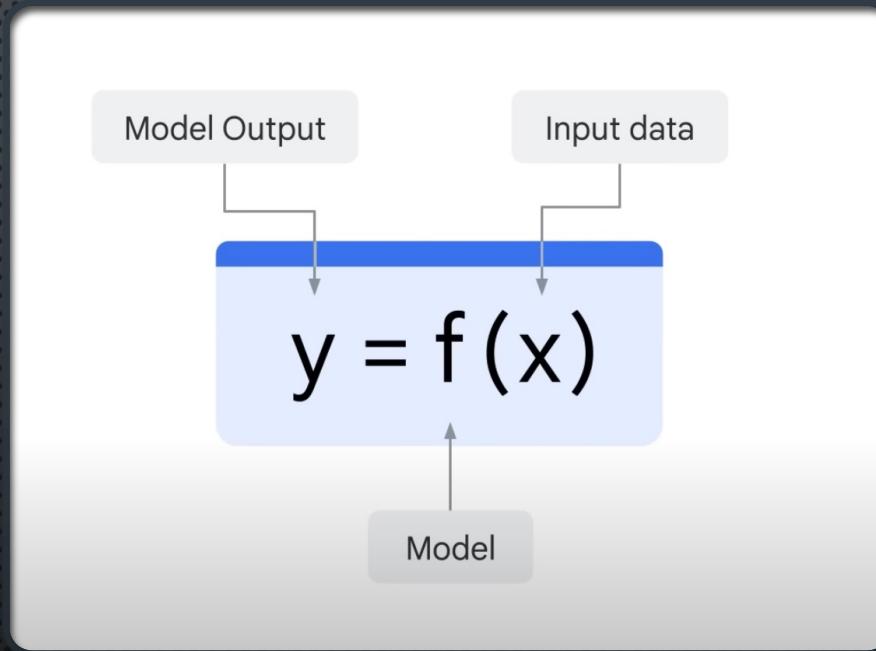
# SUPERVISED LEARNING

**Supervised learning  
implies the data is  
already labeled**

In supervised learning we are learning from past examples to predict future values.



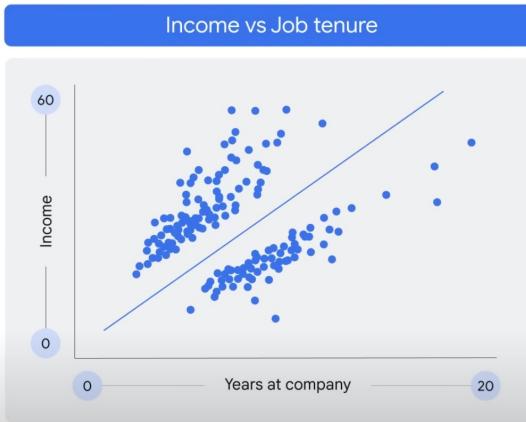
# SUPERVISED LEARNING



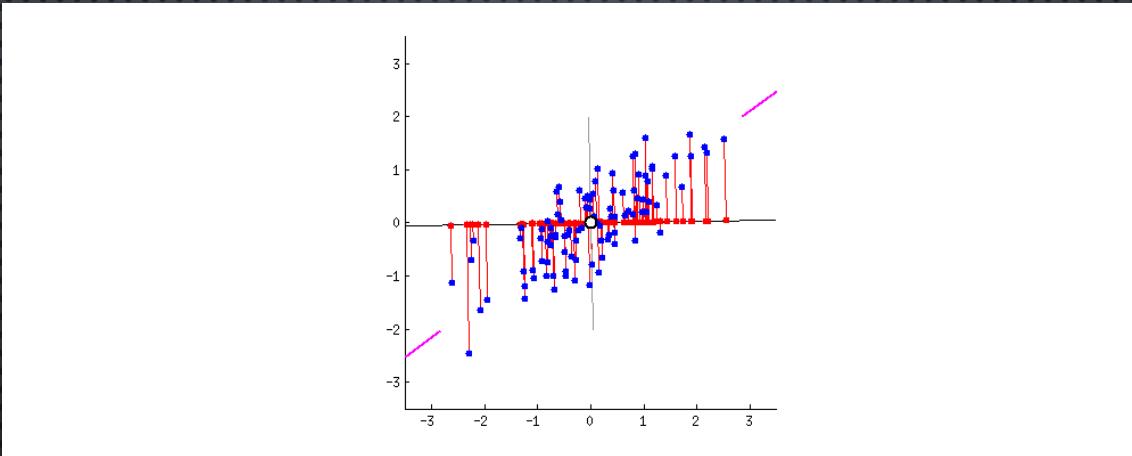
# UNSUPERVISED LEARNING

Unsupervised  
learning implies the  
data is **not labeled**

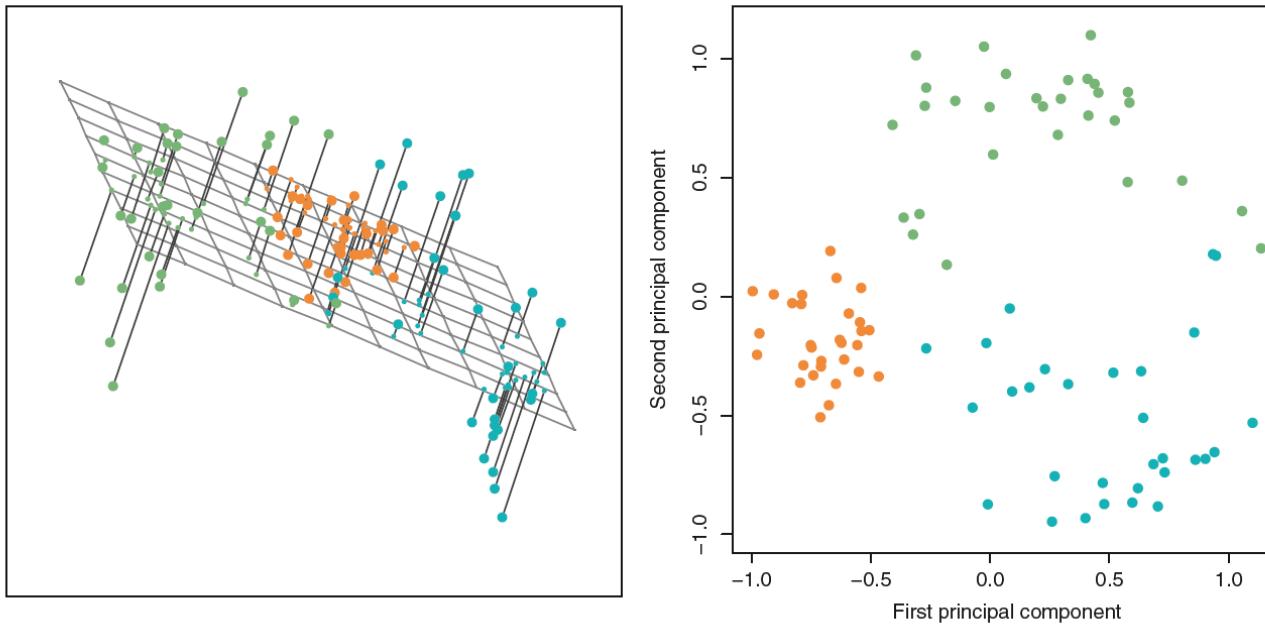
Unsupervised problems are all  
about looking at the raw data, and  
seeing if it naturally falls into groups



# UNSUPERVISED LEARNING



# UNSUPERVISED LEARNING

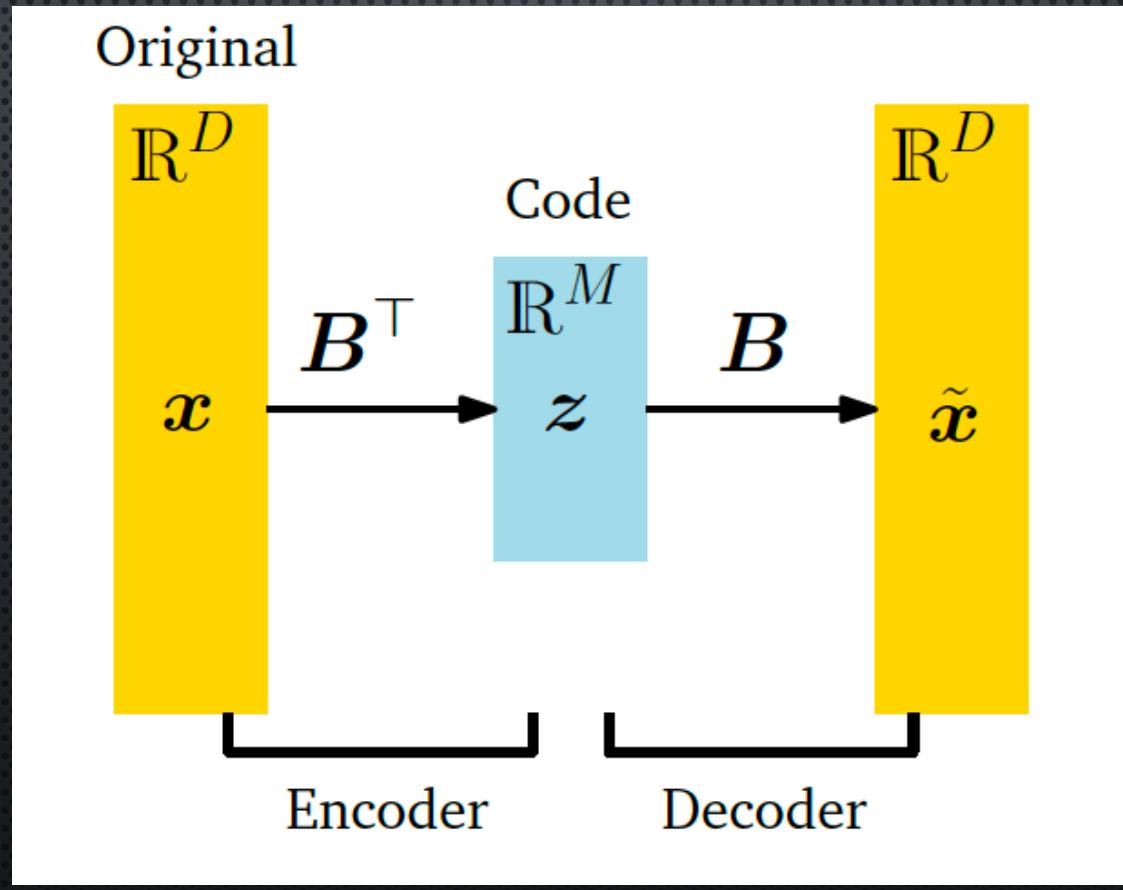


**FIGURE 10.2.** Ninety observations simulated in three dimensions. Left: the first two principal component directions span the plane that best fits the data. It minimizes the sum of squared distances from each point to the plane. Right: the first two principal component score vectors give the coordinates of the projection of the 90 observations onto the plane. The variance in the plane is maximized.

# UNSUPERVISED LEARNING

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

# IMPORTANT CONCEPT



# APPLICATIONS OF THIS IDEA

FINDING STRUCTURE IN MOVIE RATINGS

PATTERNS IN CONSUMER BEHAVIOUR

STRATIFYING PATIENTS

# ASSUMPTIONS

LINEARITY (LINEAR RELATIONSHIP BETWEEN DATA POINTS AND LOWER DIMENSIONAL REPRESENTATION)

LOSS FUNCTION/RECONSTRUCTION ERROR (SQUARED LOSS)

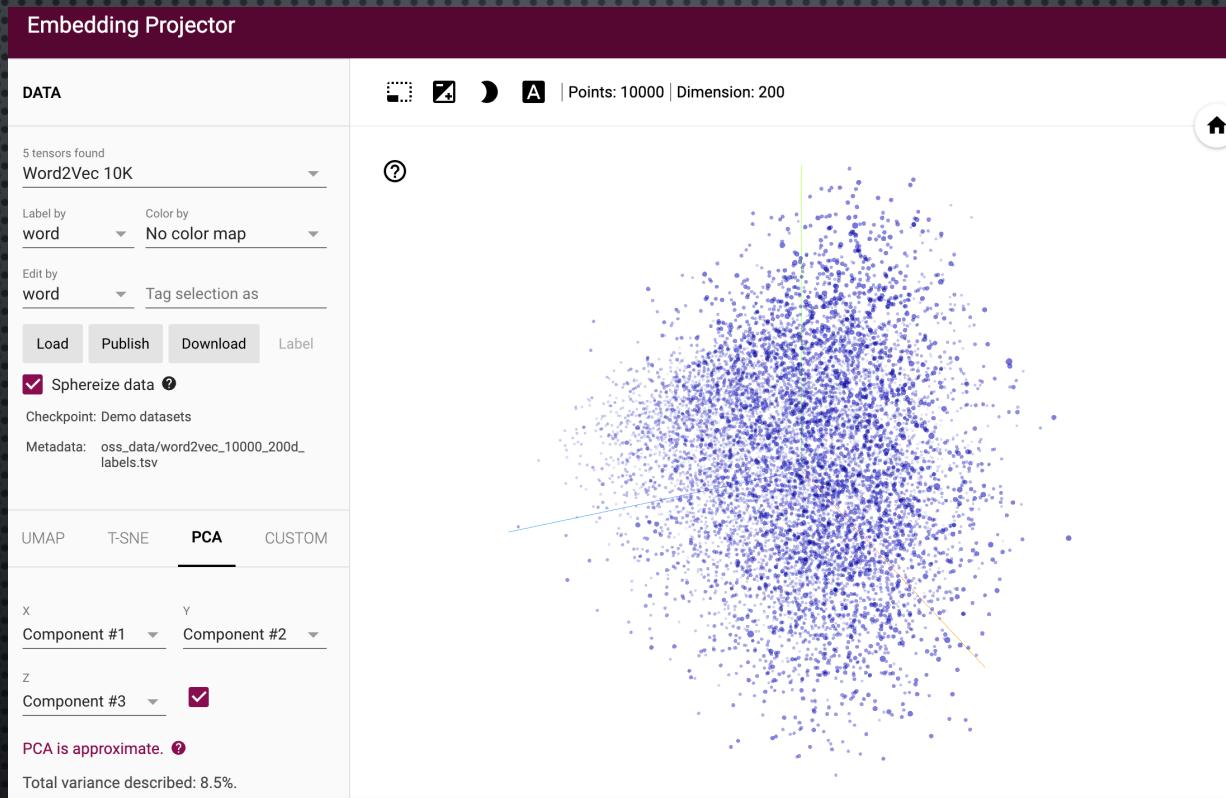
USES THE DOT PRODUCT (ONE TYPE OF INNER PRODUCT)

# GENERALIZATIONS OF THIS IDEA

TSNE

AUTOENCODER (NON-LINEAR LOSS FUNCTION)

# ACTIVITIES



[HTTP://PROJECTOR.TENSORFLOW.ORG/](http://PROJECTOR.TENSORFLOW.ORG/)

# MATERIAL

MATERIAL, CODE, EXERCISES, ACTIVITIES

[HTTPS://GITHUB.COM/NEELSOUMYA/VISUALIZATION\\_LECTURE](https://github.com/neelsoumya/visualization_lecture)

DERIVATIONS AND TECHNICAL DETAILS

[HTTPS://GITHUB.COM/NEELSOUMYA/VISUALIZATION\\_LECTURE/BLOB/MAIN/MATHEMATICS\\_DATA\\_SCIENCE.PDF](https://github.com/neelsoumya/visualization_lecture/blob/main/mathematics_data_science.pdf)