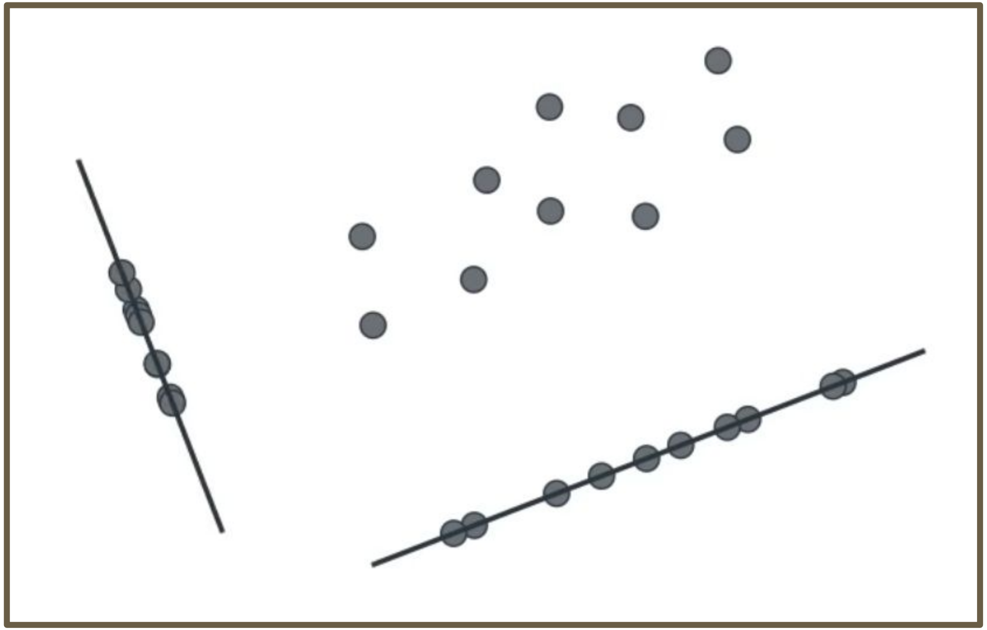
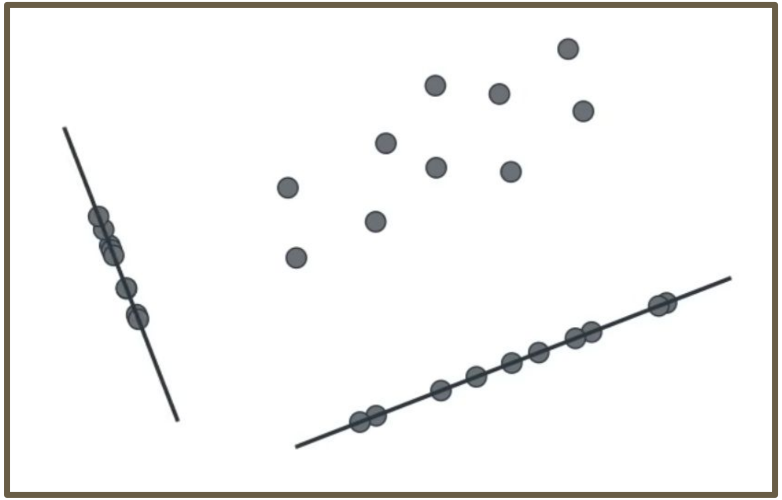
Principal Component Analysis

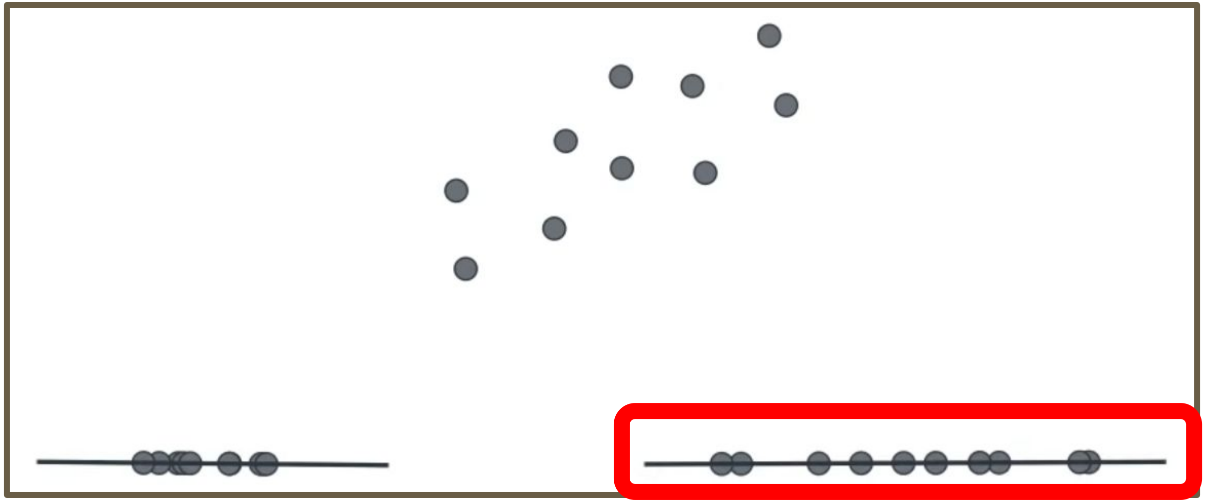
How

Source:
https://youtu.be/g-Hb26agBFg?si=ysiJ3FljeMHsLhTt | Serrano Academy

Best angle to taking a picture



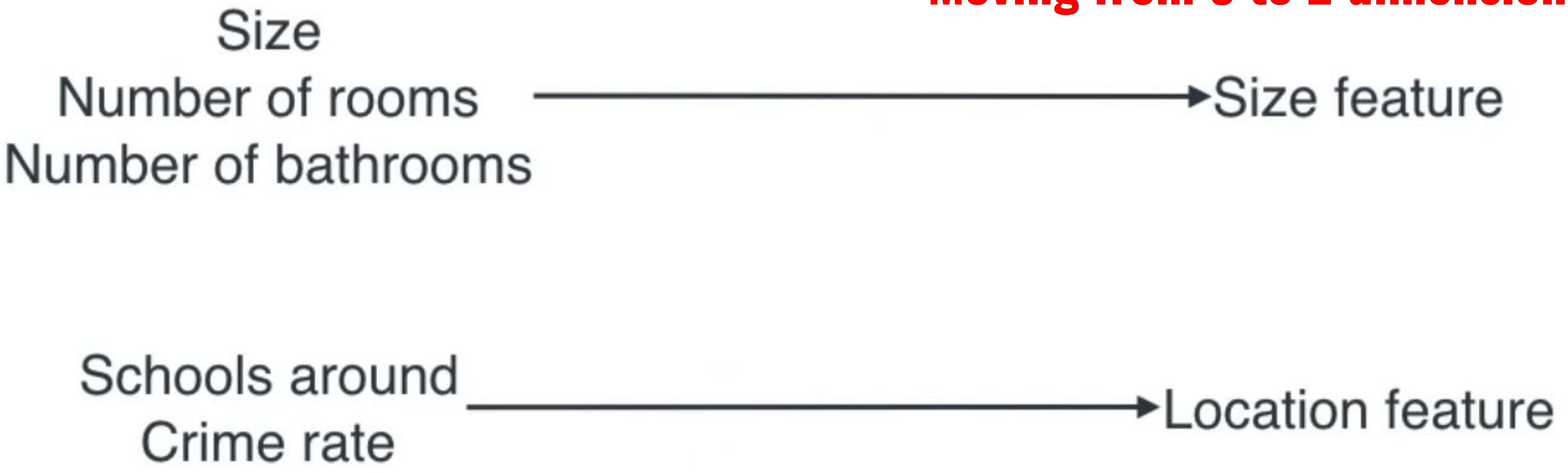




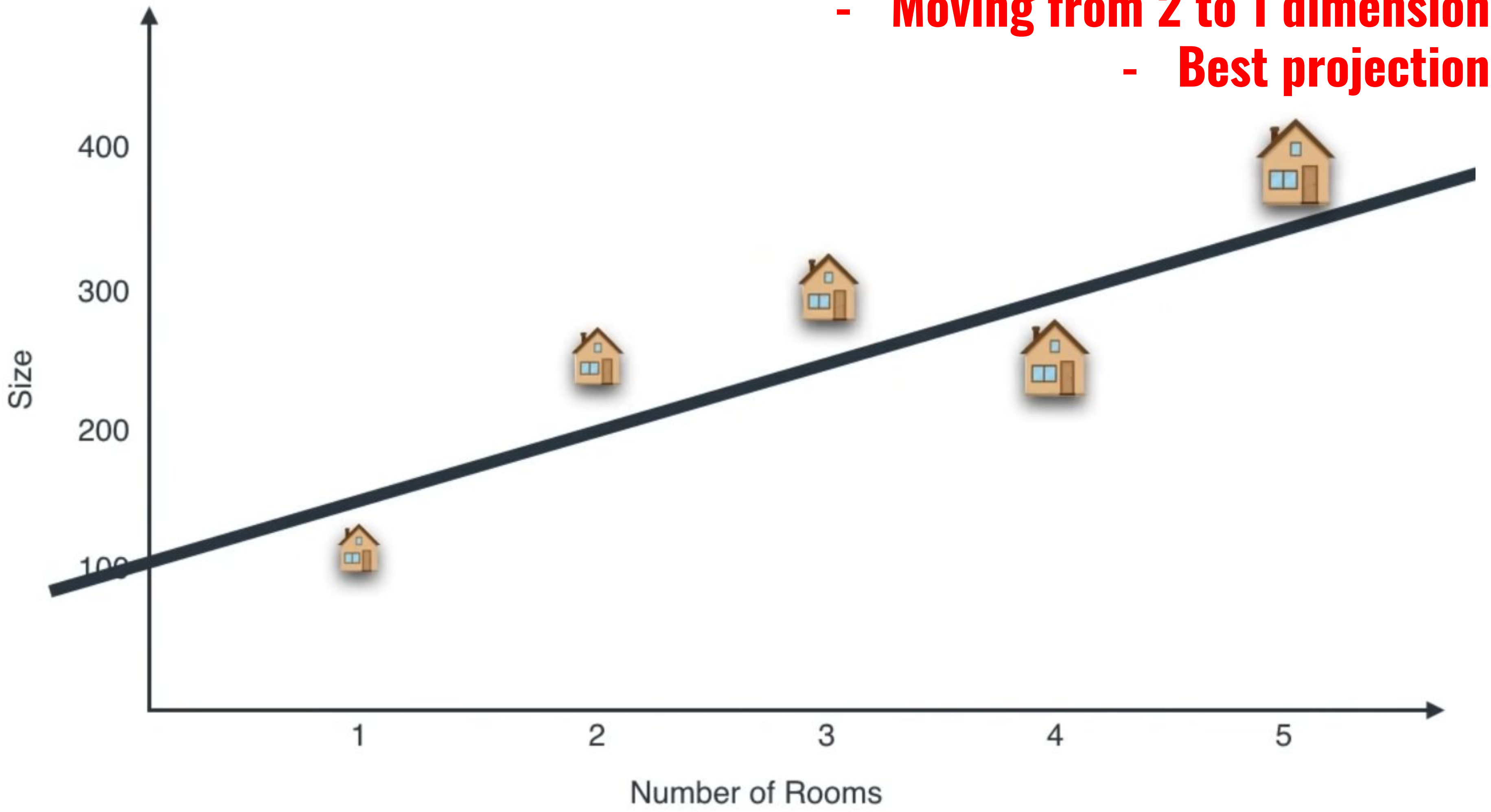
- Which projection is better?
-
- How to find that?

Why: Housing data example

- **Moving from 5 to 2 dimensions**



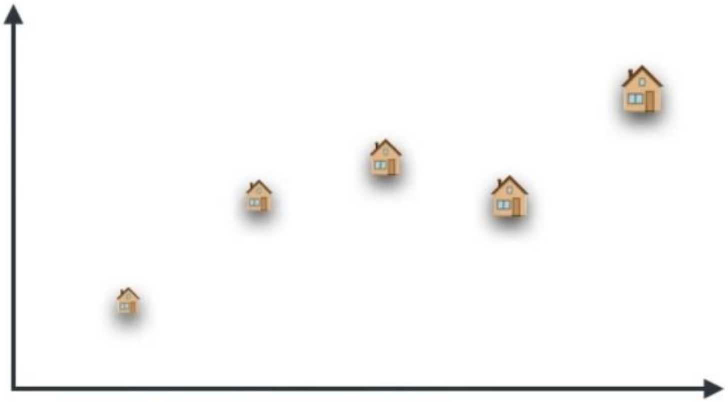
- Moving from 2 to 1 dimension
- Best projection



5 dimensions	2 dimensions
Size Number of rooms Number of bathrooms	Size feature
Schools around Crime rate	Location feature

2 dimensions

size
number of rooms

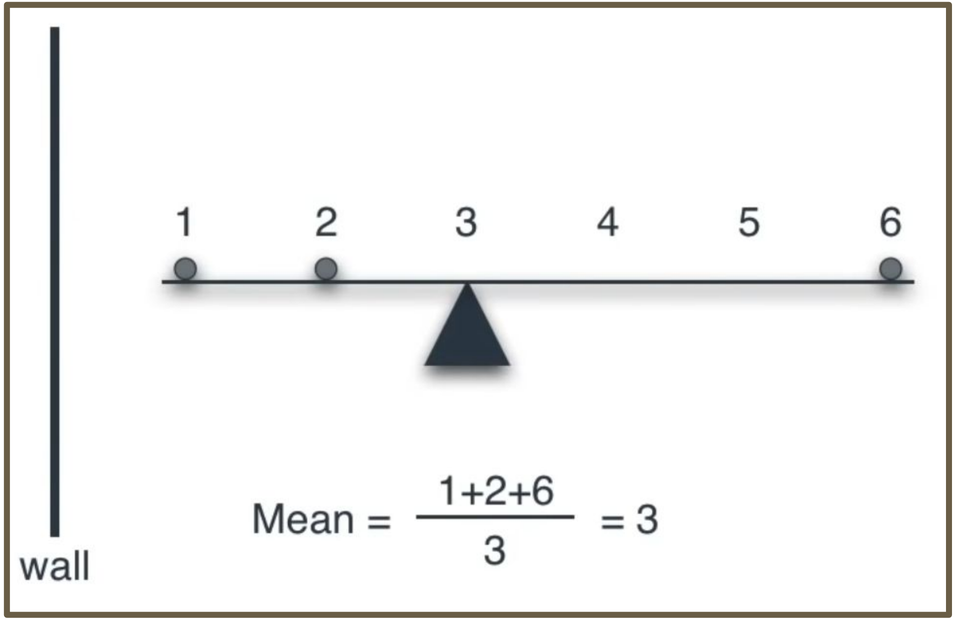


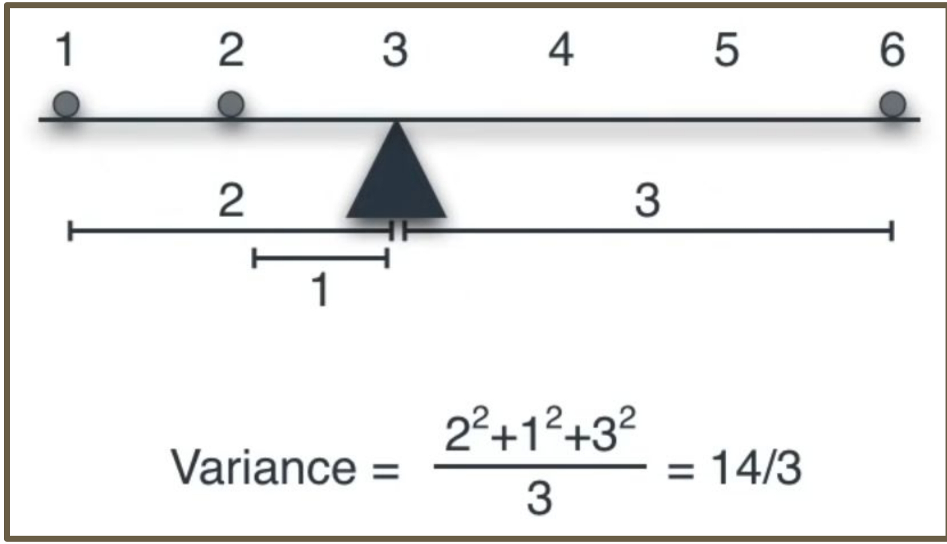
1 dimension

size feature

- **Moving from 2 to 1 dimension**
  - **Best projection**

Some Statistics - Mean vs Variance





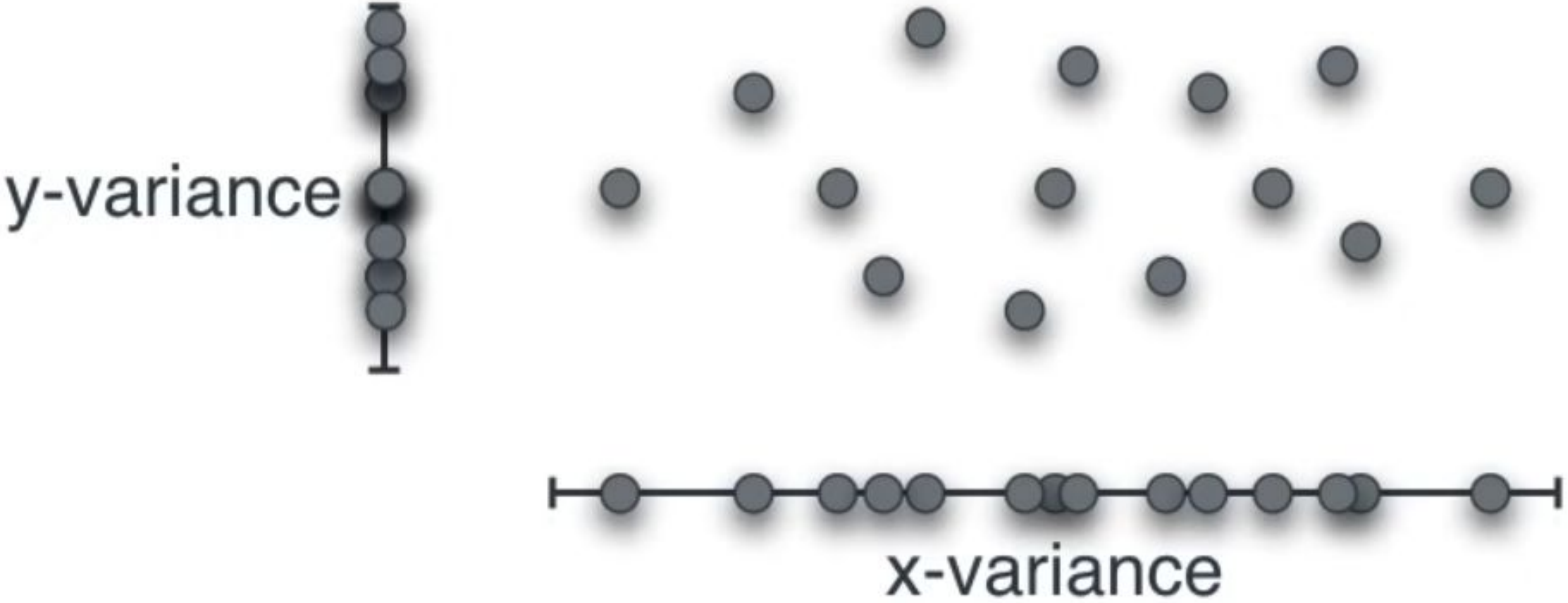
Some Statistics - Mean vs Variance

Same Mean but different variance !!

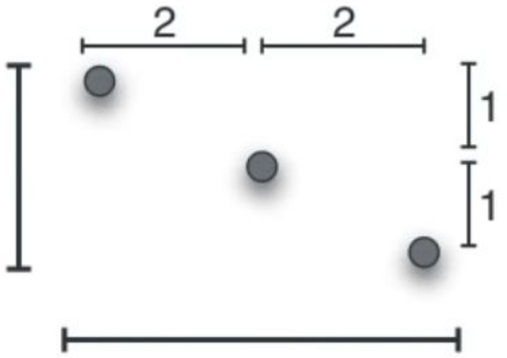
$$\text{Variance} = \frac{1^2+0^2+1^2}{3} = 2/3$$

$$\text{Variance} = \frac{5^2+0^2+5^2}{3} = 50/3$$

Some Statistics - Variance in 2D

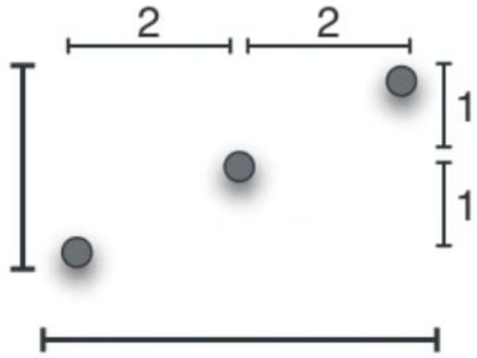


Some Statistics - Variance in 2D - Difficulties !!



Both distributions have the same X & Y variance

- **How to differentiate?**

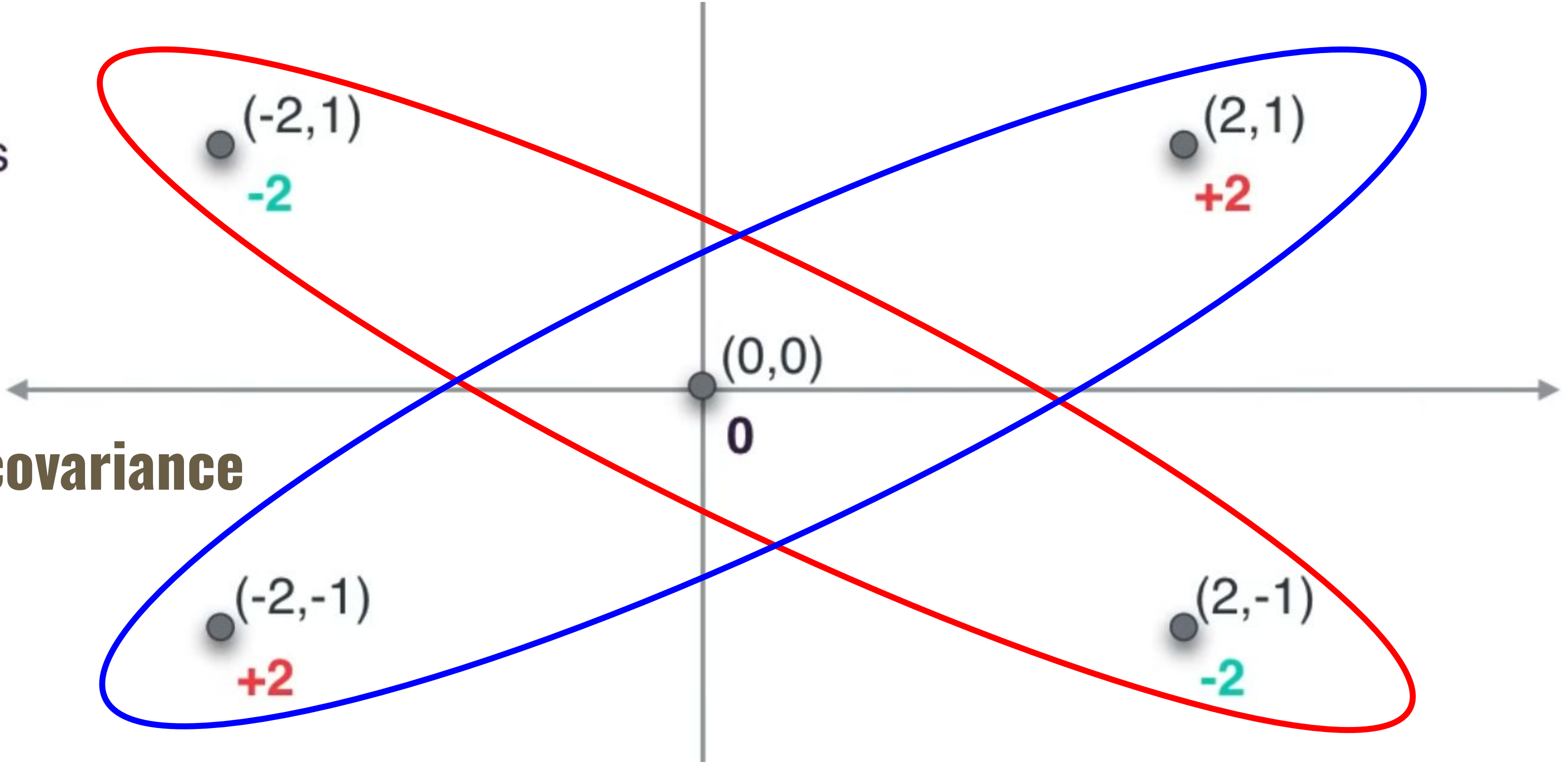


$$\text{x-variance} = \frac{2^2+0^2+2^2}{3} = 8/3$$

$$\text{y-variance} = \frac{1^2+0^2+1^2}{3} = 2/3$$

Some Statistics - Covariance

Product of coordinates

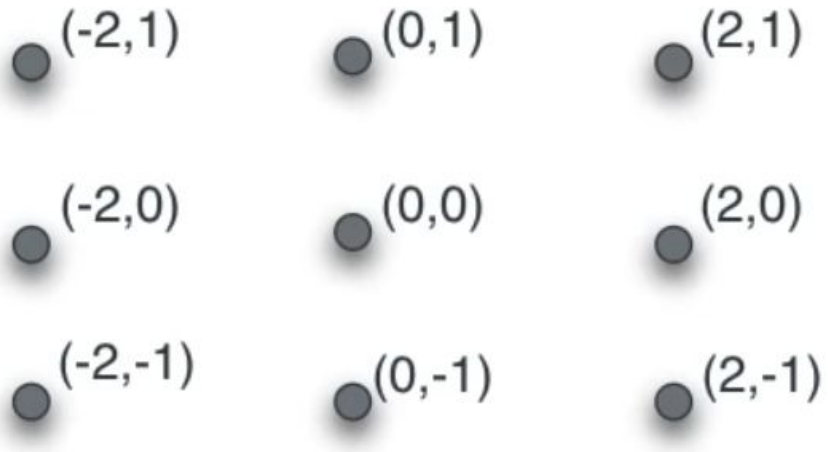


+ve & -ve covariance

$$\text{covariance} = \frac{(-2) + 0 + (-2)}{3} = -4/3$$

$$\text{covariance} = \frac{2 + 0 + 2}{3} = 4/3$$

Some Statistics - Zero Covariance



$$\text{covariance} = \frac{-2 + 0 + 2 + 0 + 0 + 0 + 2 + 0 + -2}{9} = 0$$

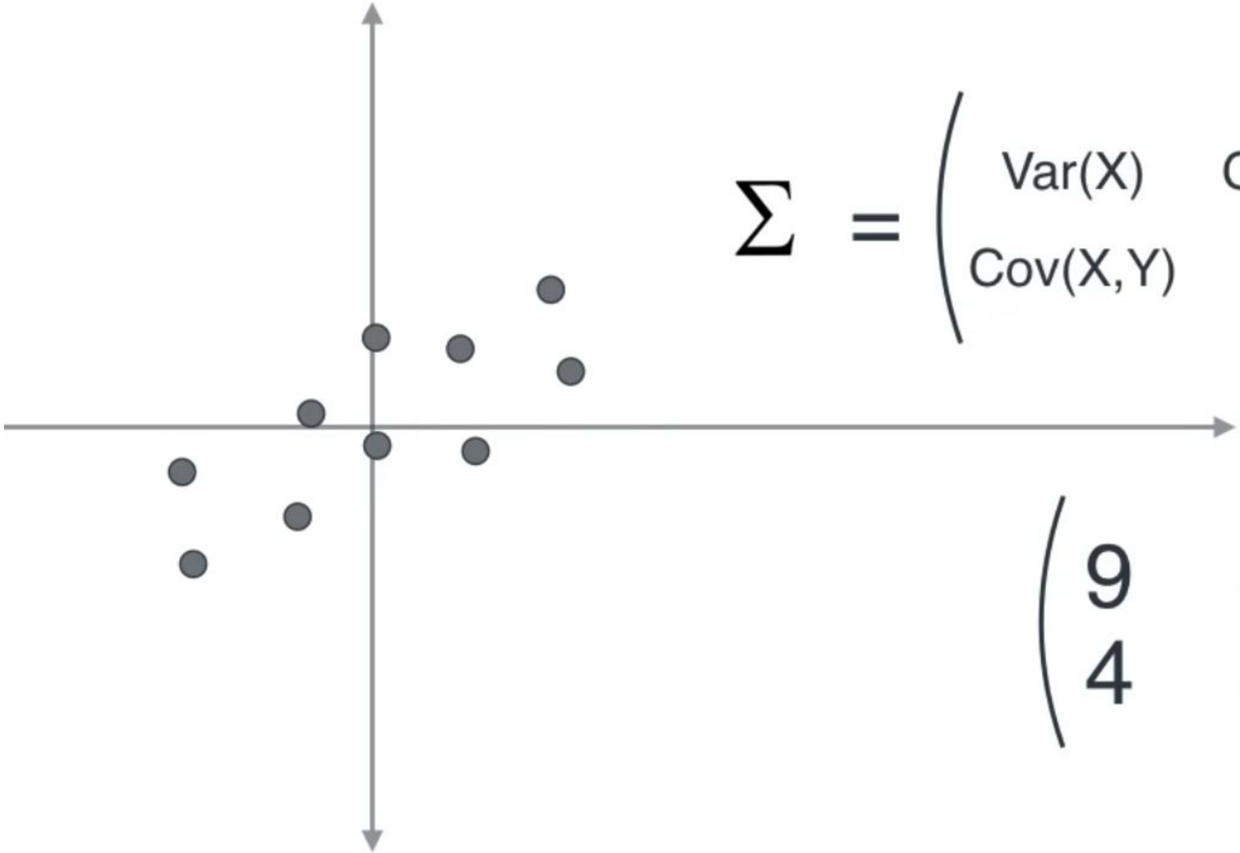
Some Statistics - Covariance - Summary

negative covariance

covariance zero (or very small)

positive covariance

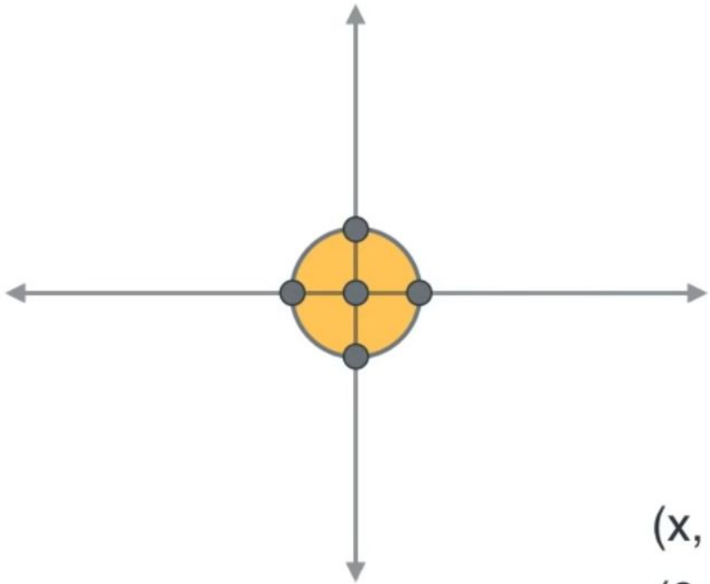
Some Statistics - Covariance Matrix



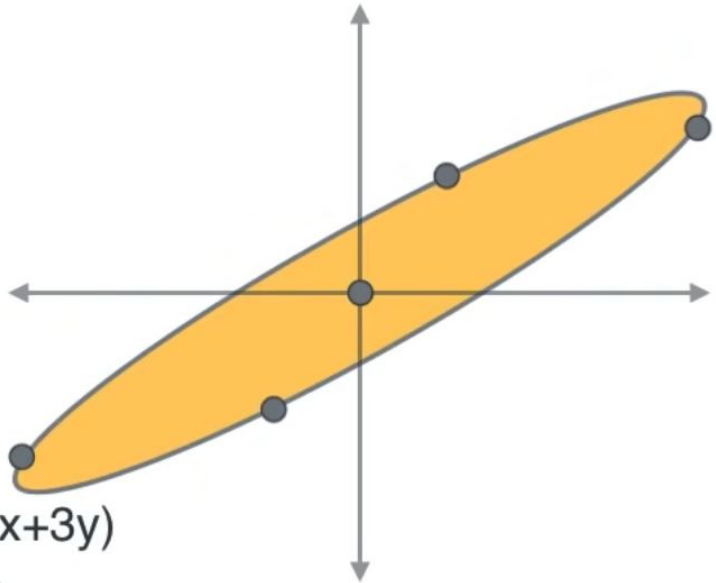
$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}$$

$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

Some Statistics - Linear Transformations



$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} \longrightarrow$$

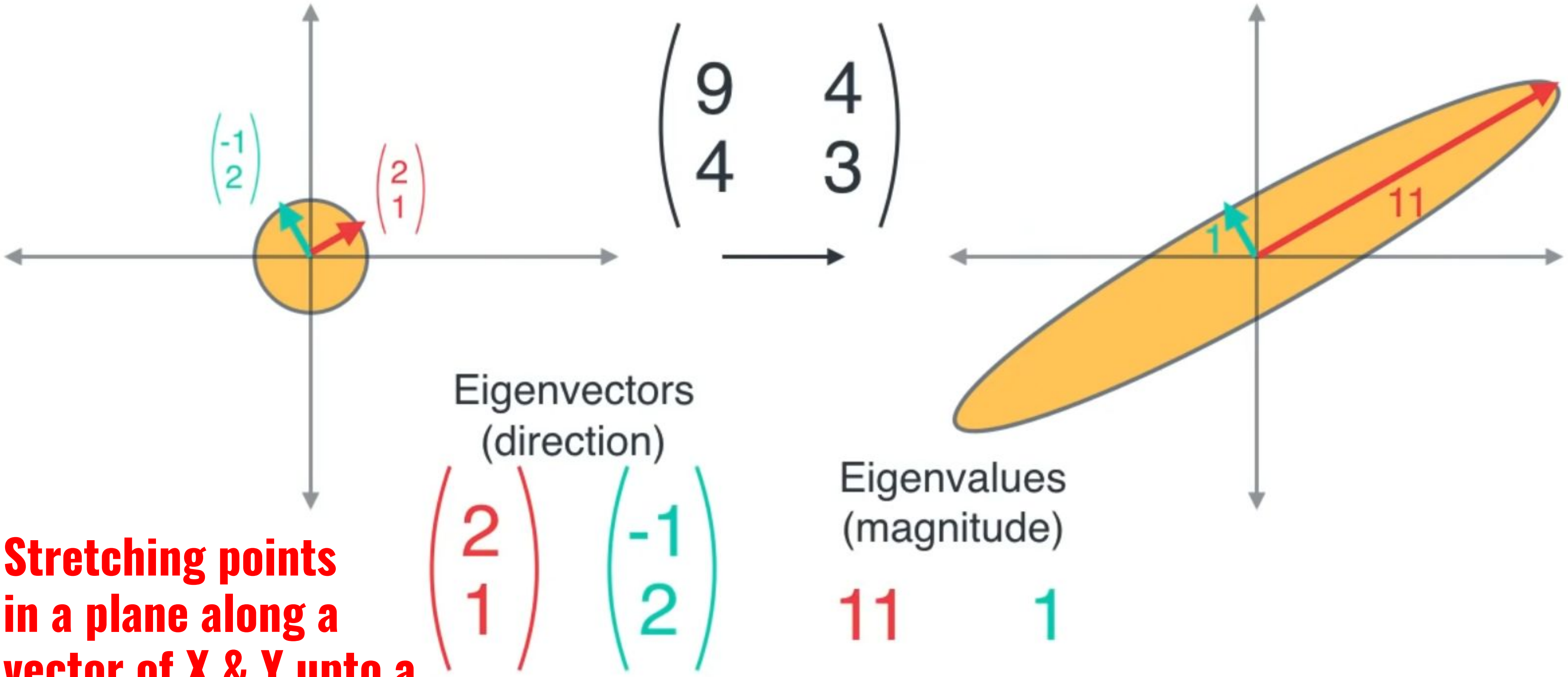


$(x, y) \longrightarrow (9x+4y,\ 4x+3y)$

(0,0)	(0,0)
(1,0)	(9,4)
(0,1)	(4,3)
(-1,0)	(-9,-4)
(0,-1)	(-4,-3)

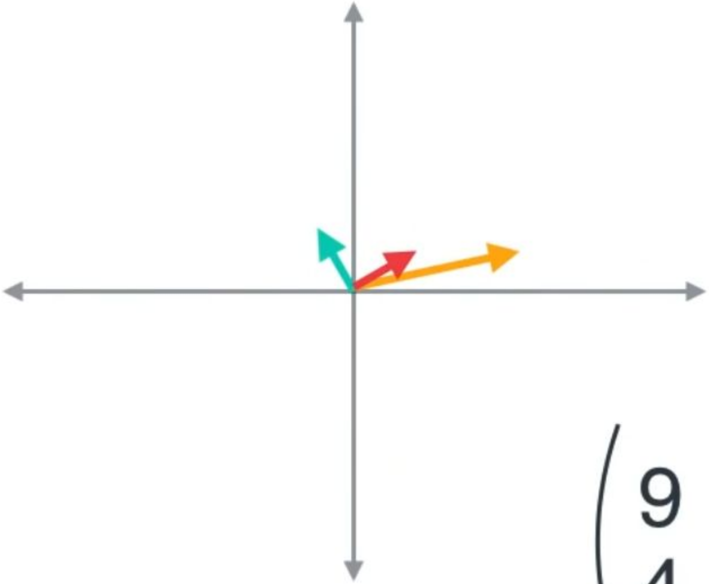
- **Stretching points in a plane along a vector of X & Y**

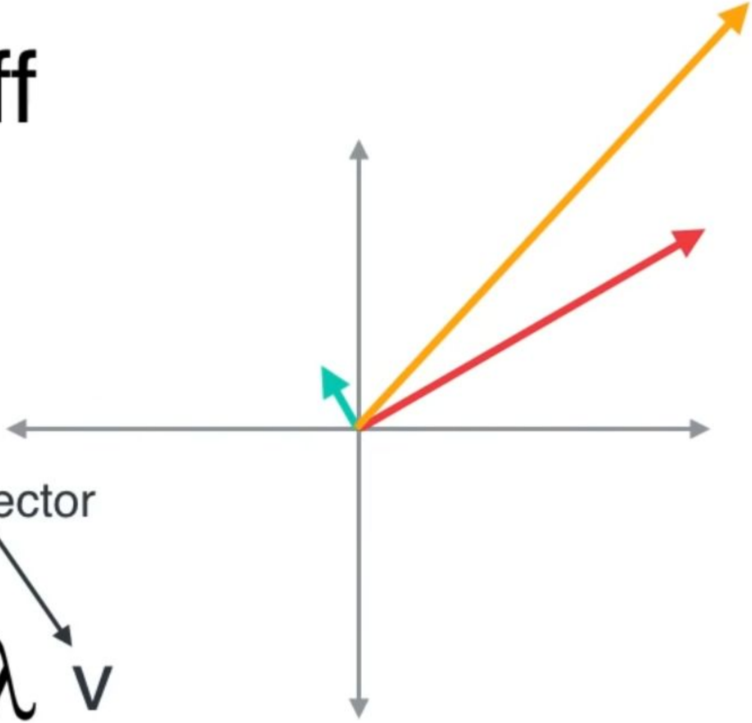
Some Statistics - Linear Transformations



Stretching points in a plane along a vector of X & Y upto a scalar magnitude

Eigenstuff



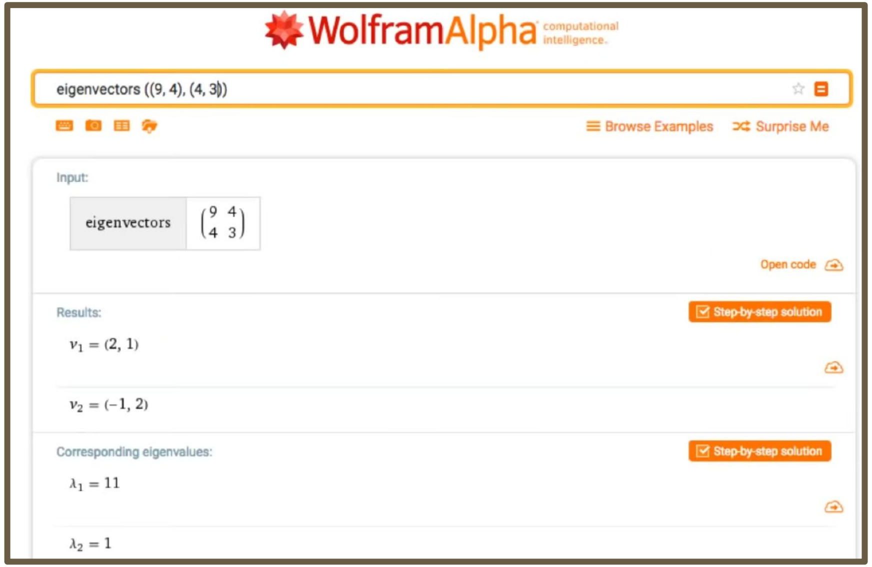


Eigenvector

$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} v = \lambda \, v$$

Eigenvalue

Finding Eigenvalues ?



$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

Characteristic Polynomial

$$\begin{vmatrix} x-9 & -4 \\ -4 & x-3 \end{vmatrix} = (x-9)(x-3) - (-4)(-4) = x^2 - 12x + 11$$

$$= (x-11)(x-1)$$

Eigenvalues 11 and 1

Finding Eigenvectors ?

$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = 11\begin{pmatrix} u \\ v \end{pmatrix}$$

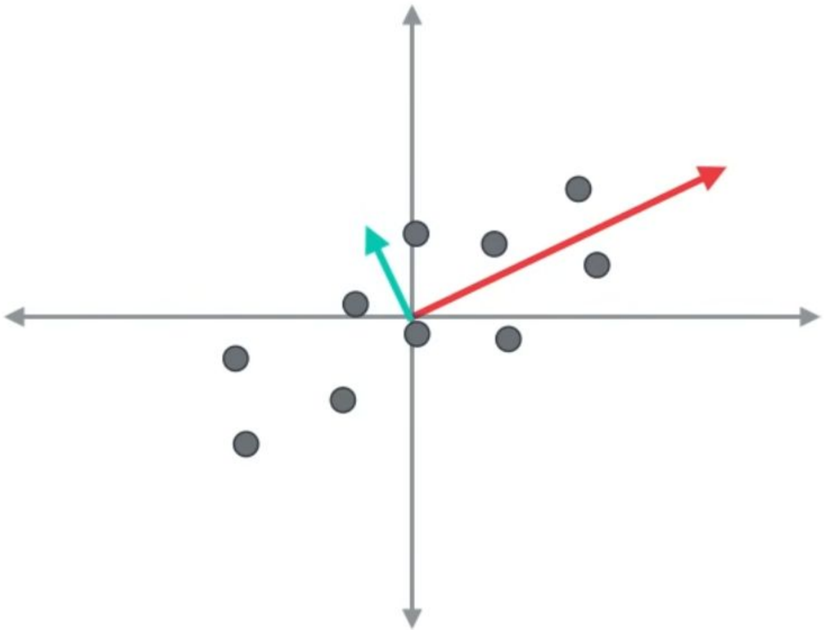
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = 1\begin{pmatrix} u \\ v \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

PCA

\# It tells us - what direction the distribution is spread and how much?
\# Note - G & R vectors are perpendicular for symmetric cov-matrix
\# Which vector (PCA) is more important? Ans: R

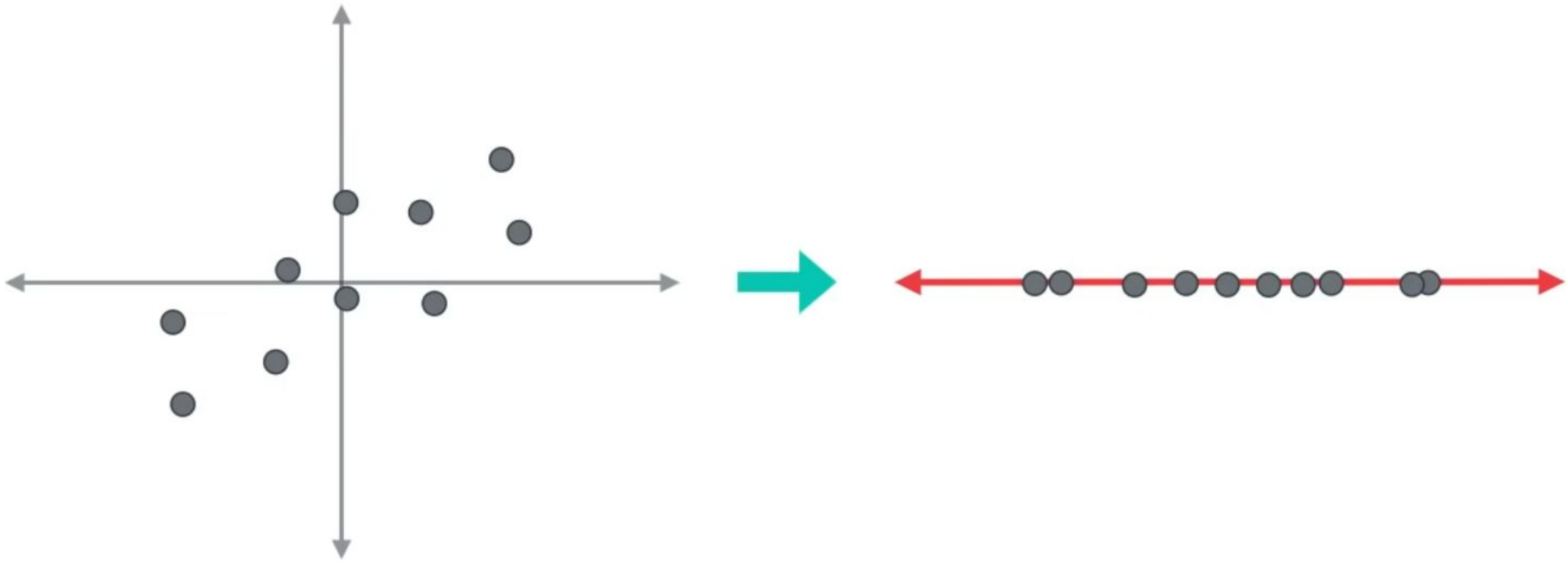


$$\Sigma = \begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$$

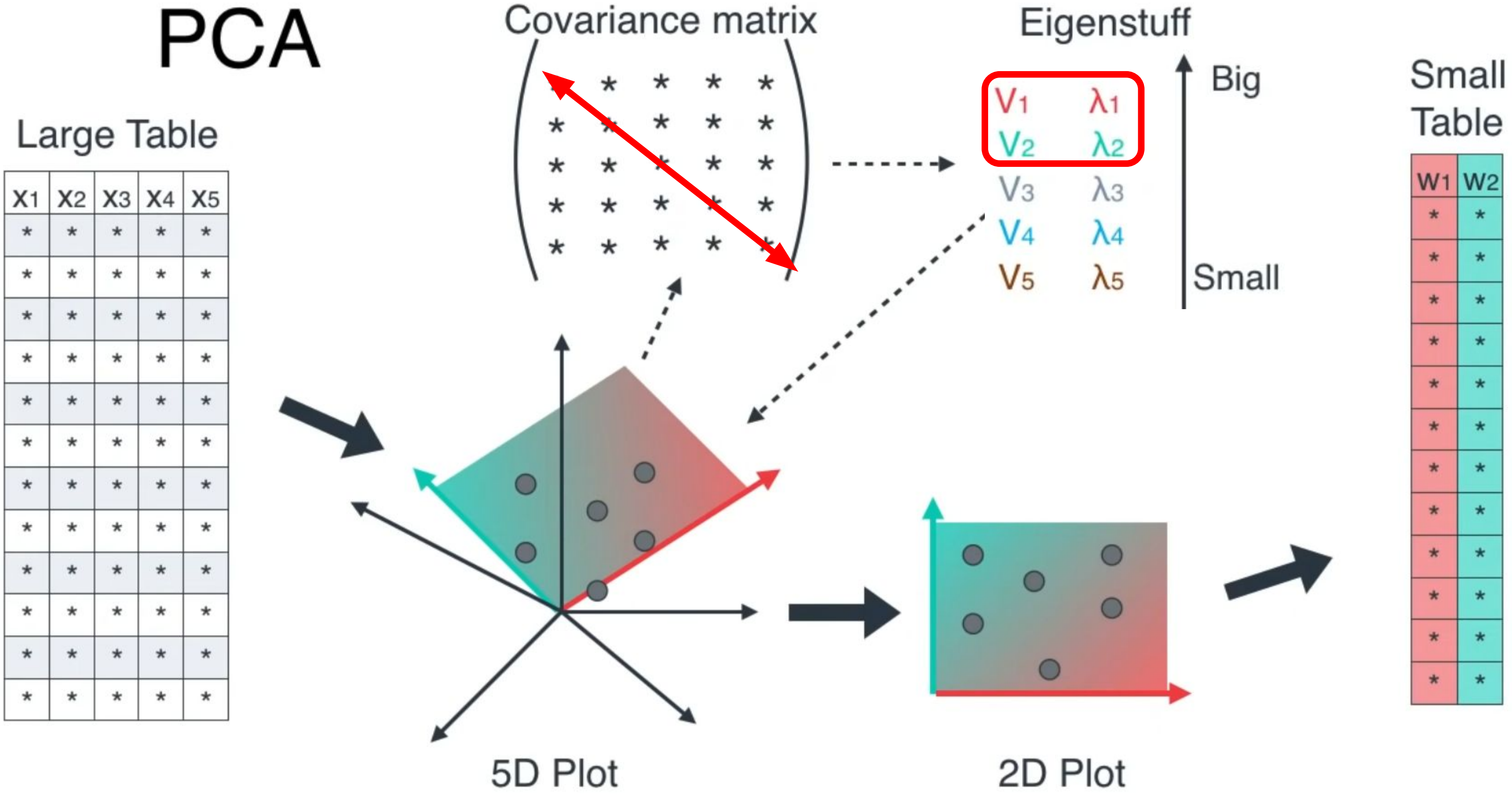
$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$ Eigenvectors (direction)

$$11 \quad 1$$ Eigenvalues (magnitude)

PCA-1



PCA


Covariance matrix
Eigenstuff
Large Table
X1 X2 X3 X4 X5
V1 λ1
V2 λ2
V3 λ3
V4 λ4
V5 λ5
Big
Small
Small Table
W1 W2
5D Plot
2D Plot