

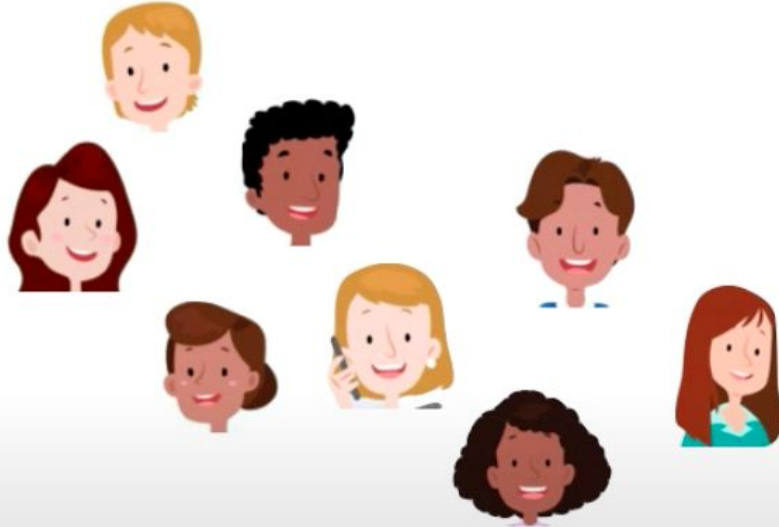
---

# PCA- Análisis de Componentes Principales

— Silvana Armas  
Guillermo Rivera  
Alejandra Salinas —

---

# Tomando una foto



# Tomando una foto



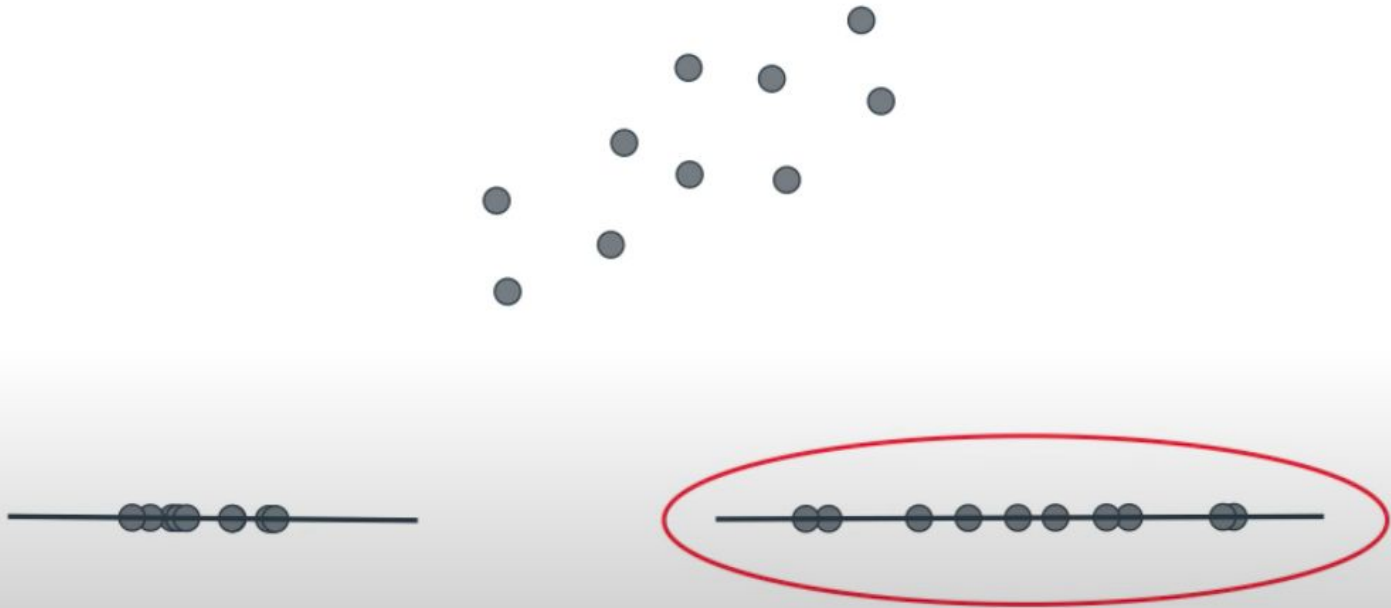
# Reduccion de dimensionalidad



# Reduccion de dimensionalidad



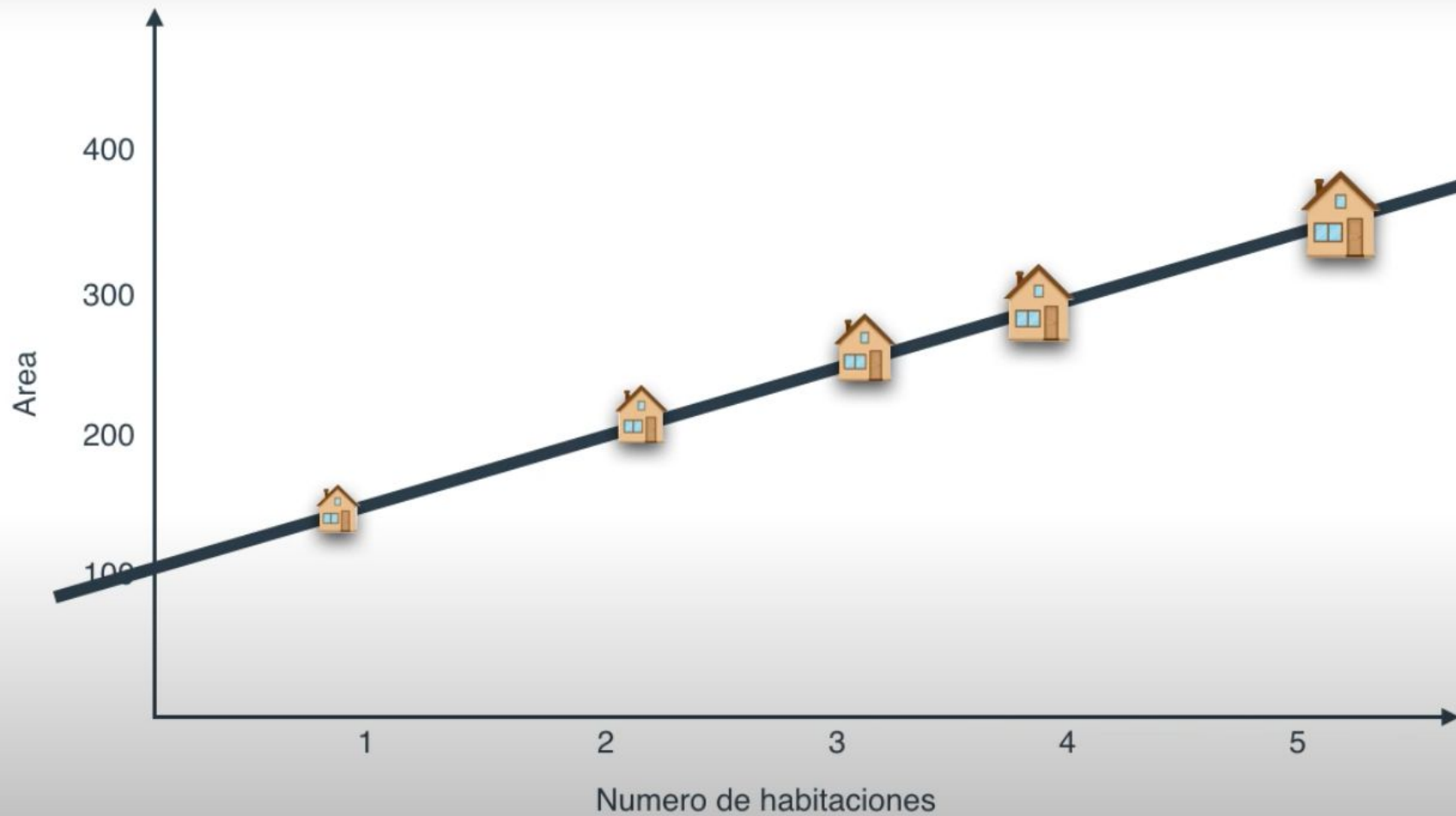
# Reduccion de dimensionalidad









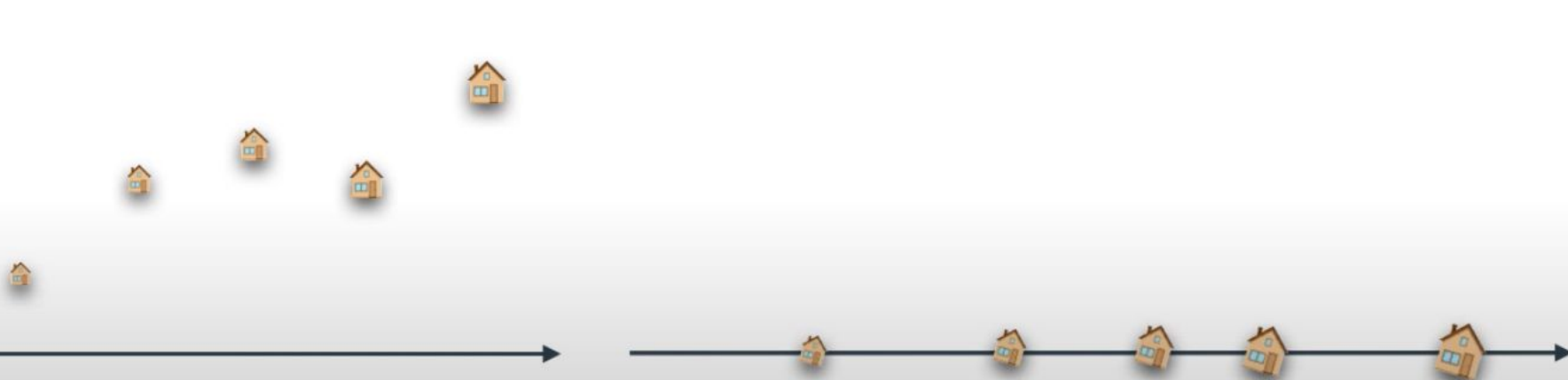


2 dimensiones

1 dimension

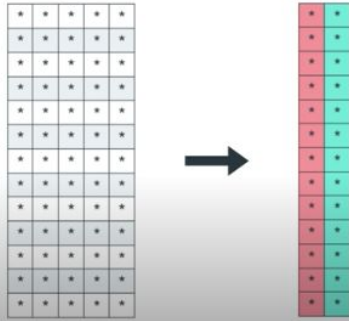
area  
numero de habitaciones

tamaño



# Introducción

## Reduccion de dimensionalidad



## Datos de casas

Area  
Numero de habitaciones  
Numero de baños  
Escuelas cercanas  
Crimen en el area

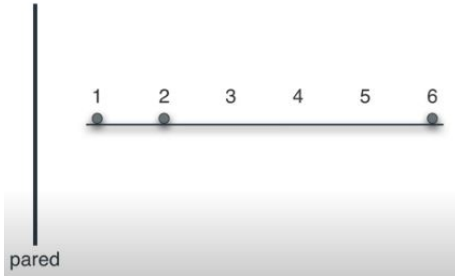
## Datos de casas

Area  
Numero de habitaciones —————→ Tamaño  
Numero de baños

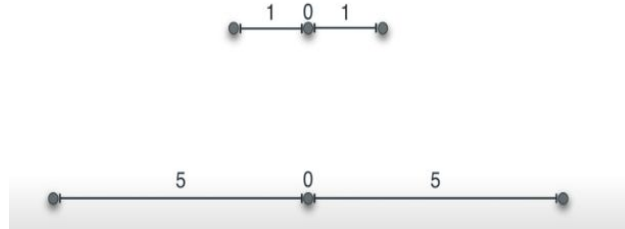
Escuelas cercanas —————→ Ubicación  
Crimen en el area

# Funcionamiento

Promedio



Varianza



Covarianza



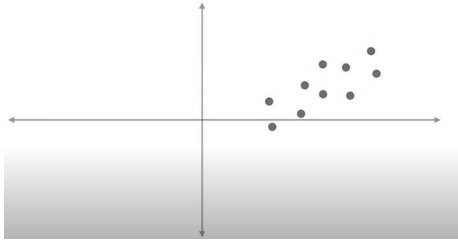
covarianza  
negativa



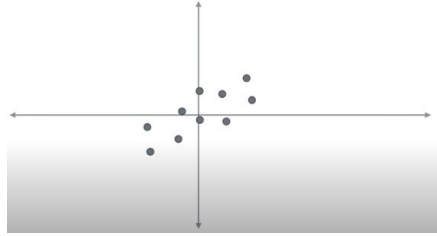
covarianza cero  
(o muy pequeña)



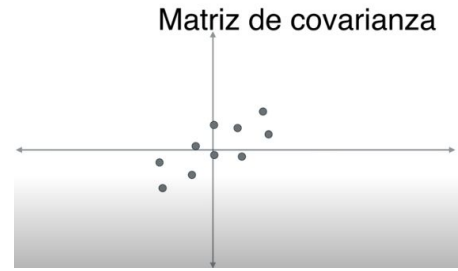
covarianza  
positiva



Se tiene el conjunto de datos.



Encontrar el centro de masa.  
Promedio de las coordenadas x y y.

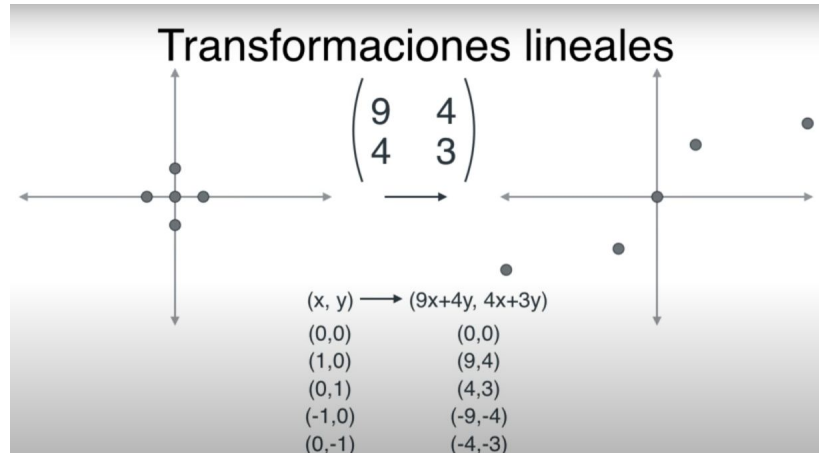


Matriz de covarianza

Qué tan espacidos están los datos,  
hacia x, y, positivo o negativo.

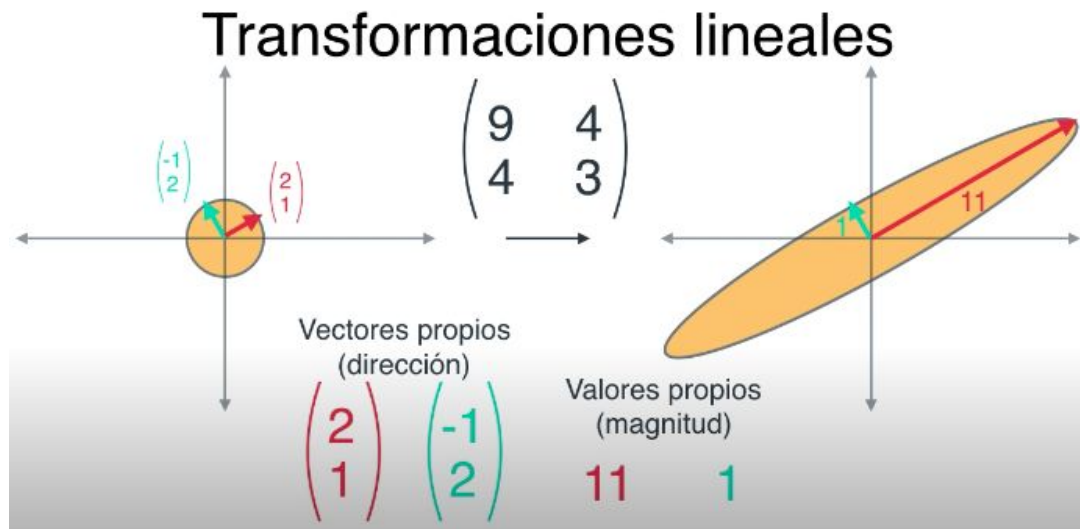
$$\Sigma = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X,Y) \\ \text{Cov}(X,Y) & \text{Var}(Y) \end{pmatrix}$$

Realiza la matriz de covarianza.



Manda los número de la matriz a coordenadas del plano.

# Eigenvalues & Eigenvectors



Explica la transformación lineal, de estirar un plano en dos direcciones perpendiculares.

# Ventajas

Permite simplificar y manejar datos que tienen varias dimensiones.

# Desventajas

Existe un porcentaje de varianza que no va a ser explicado por el modelo.

Es limitado cuando los datos están expresados en distintas escalas.

# Problemas en los que se aplica

## Ejemplo PCA aplicado a genómica

El siguiente es un ejemplo de como PCA puede emplearse para encontrar patrones cuando se dispone de muchos de predictores.

*Supóngase un equipo de investigación que se encarga de clasificar los tumores de pacientes en 3 subtipos. Dependiendo del subtipo, el paciente recibe una medicación diferente. El proceso de caracterización se hace mediante tinciones y observaciones al microscopio. Este proceso es muy laborioso y lento, lo que incrementa mucho el tiempo de respuesta de los médicos. Nuevos estudios apuntan a que cuantificando la expresión de un grupo de 9 genes se podría clasificar los tumores con una alta precisión. Se quiere determinar si tal patrón existe dentro de los datos.*



# Referencias

<https://www.youtube.com/watch?v=AniiwysJ-2Y>

[https://www.cienciadedatos.net/documentos/35\\_principal\\_component\\_analysis#Ejemplo\\_PCA\\_aplicado\\_a\\_genómica](https://www.cienciadedatos.net/documentos/35_principal_component_analysis#Ejemplo_PCA_aplicado_a_genómica)

<https://empresas.blogthinkbig.com/python-para-todos-que-es-el-pca/>