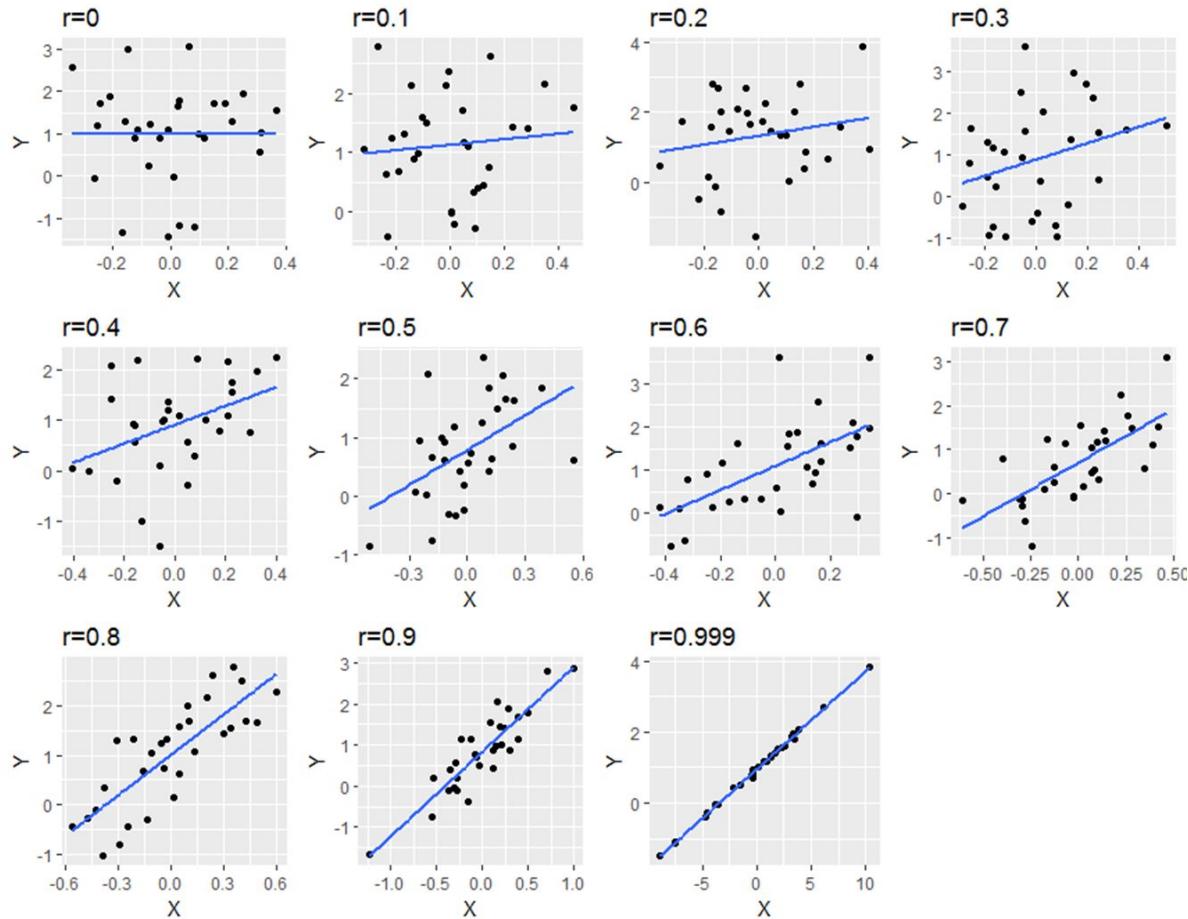
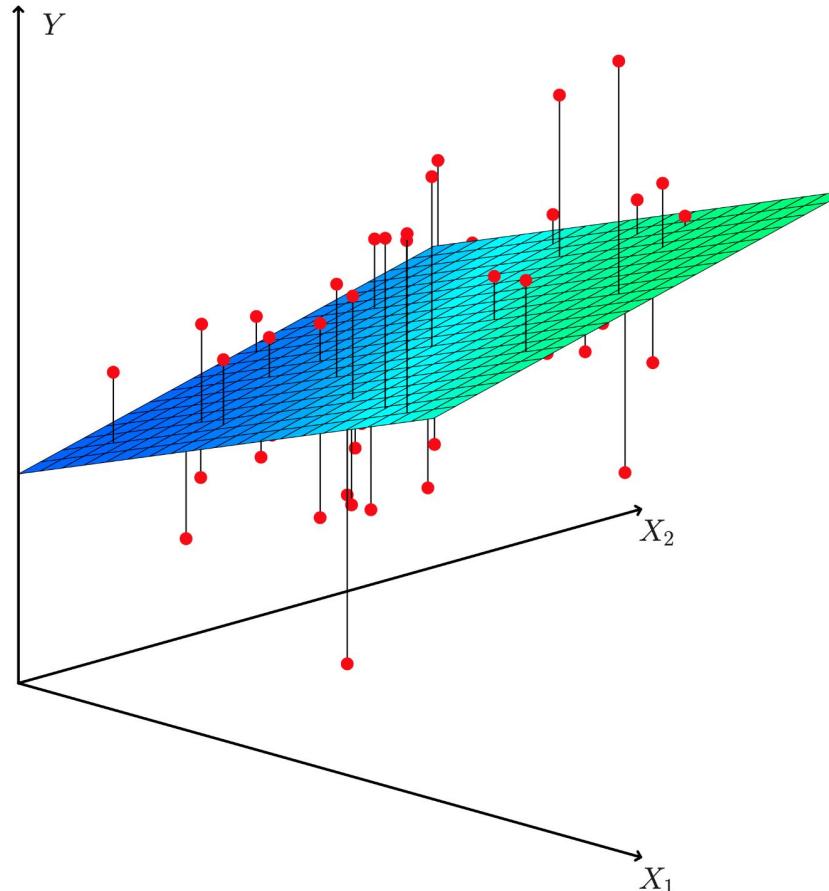
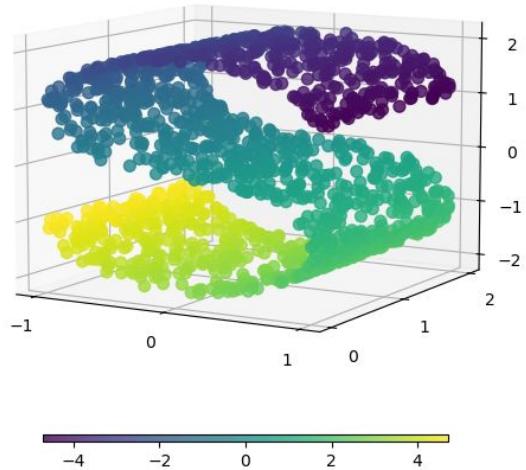


Reducción de dimensionalidad

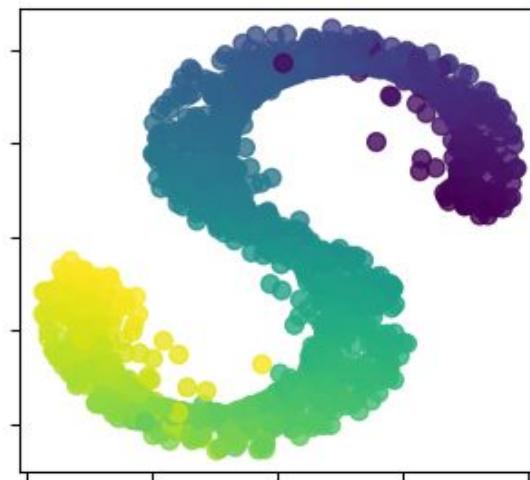




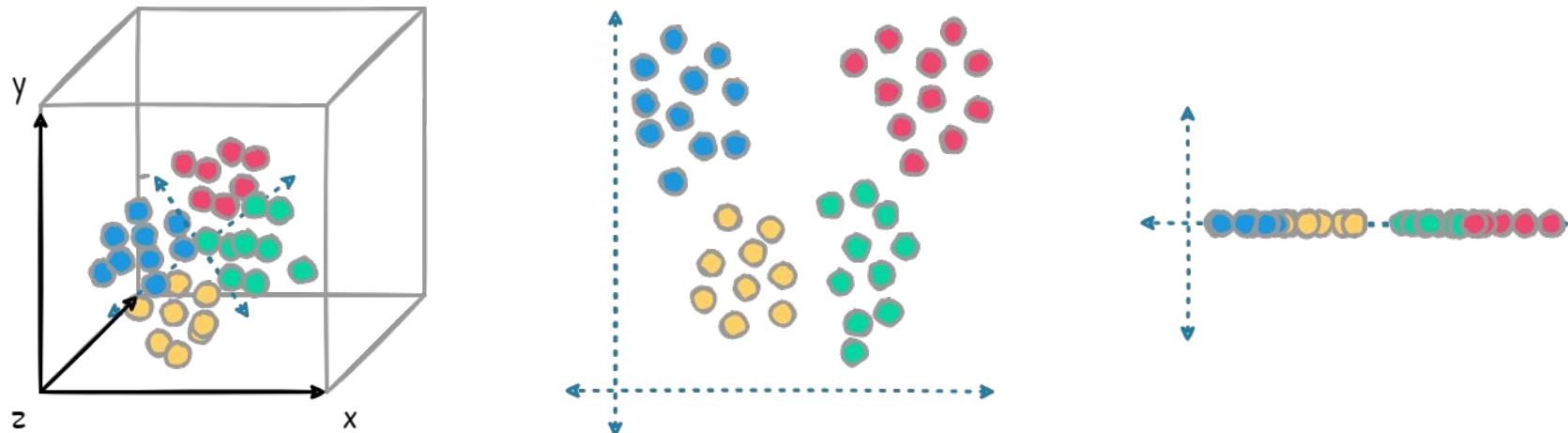
Original S-curve samples



Multidimensional scaling



Dimensionality Reduction



¿Existe alguna forma correcta de “bajar” la dimensión?

Trazos a continuación en uno, dos y tres dimensiones.

Preprocesamiento de los datos

- Evitamos que ciertas características tengan mayor relevancia
- Mayor rendimiento en distintos algoritmos
- Estabilidad numérica

$$z = \frac{x - \mu}{\sigma}$$

Dos algoritmos para escalar datos: standard scaler

$$z = \frac{x - \mu}{\sigma}$$

El centrado y el escalado se realizan de forma independiente en cada característica calculando las estadísticas relevantes en las muestras del conjunto de entrenamiento.

Este escalador es sensible a valores atípicos.

MinMaxScaler

Transformar las características escalando cada una a su rango correspondiente.

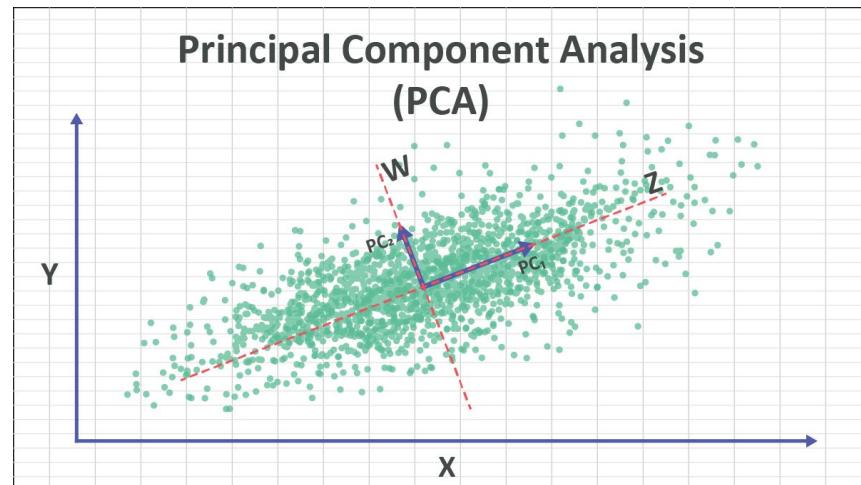
Esta transformación se utiliza a menudo como alternativa al escalamiento con media cero y varianza unitaria.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

¿Después del escalado... que hacemos?

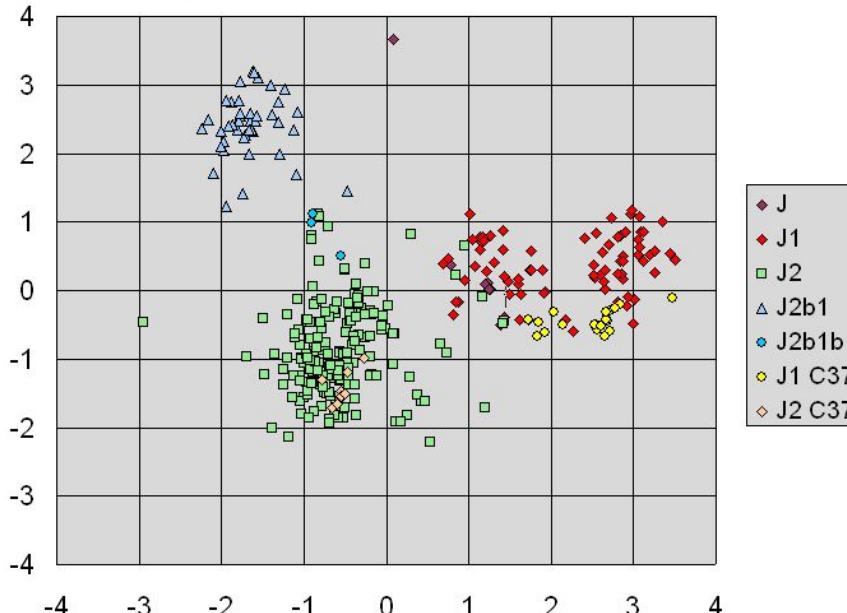
Existen varios métodos para reducir la dimensión de un conjunto de datos, entre ellos:

- PCA
- Kernel PCA
- t-SNE
- Isomap
- UMAP
- etc..



¿Para que sirve esto?

Haplogroup J - 37 STRs



Aproximación a PCA

