

Intuición sobre Análisis de Componentes Principales (PCA)

Intuición sobre PCA

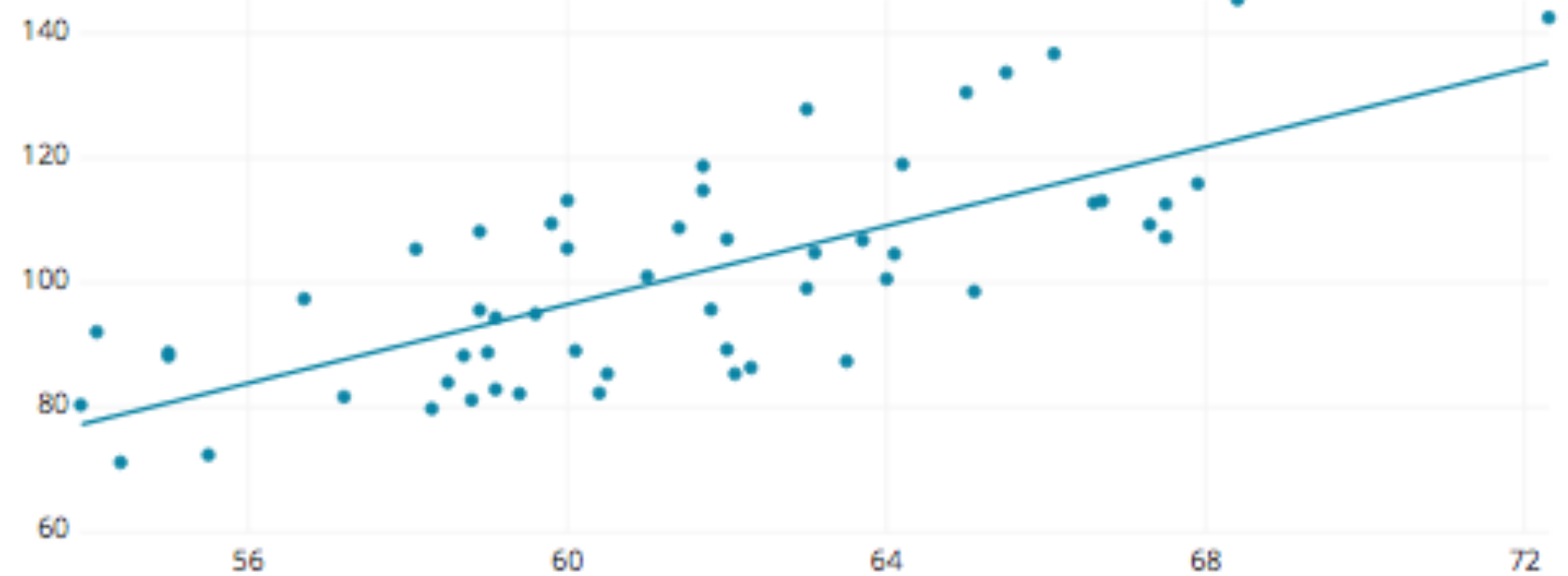
- PCA es una técnica estadística no supervisada que se utiliza para examinar las interpelaciones entre variables de un conjunto para ver si hay alguna estructura subyacente.
- También es conocida como un **Análisis de Factores** general

Intuición sobre PCA

- Regresión determina una línea que mejor se ajuste a un conjunto de datos
- PCA determina varias líneas ortogonales que mejor se ajustan a los datos
- Ortogonal quiere decir “a ángulos rectos”
 - En realidad las líneas son perpendiculares entre si en el espacio n-dimensional
- El espacio n-dimensional es el espacio de las variables de las observaciones
 - Hay tantas dimensiones como hay variables

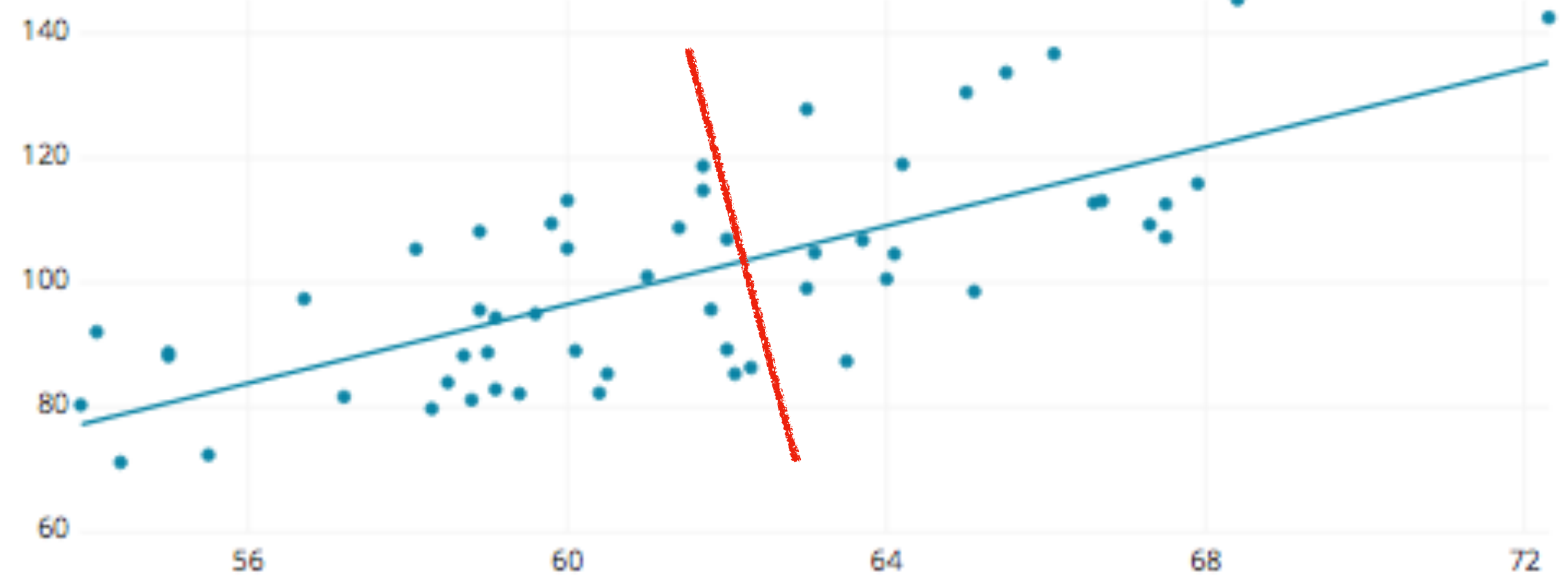
Intuición sobre PCA

- Tomemos unos datos gratinados sobre dos ejes que representan las características x e y .



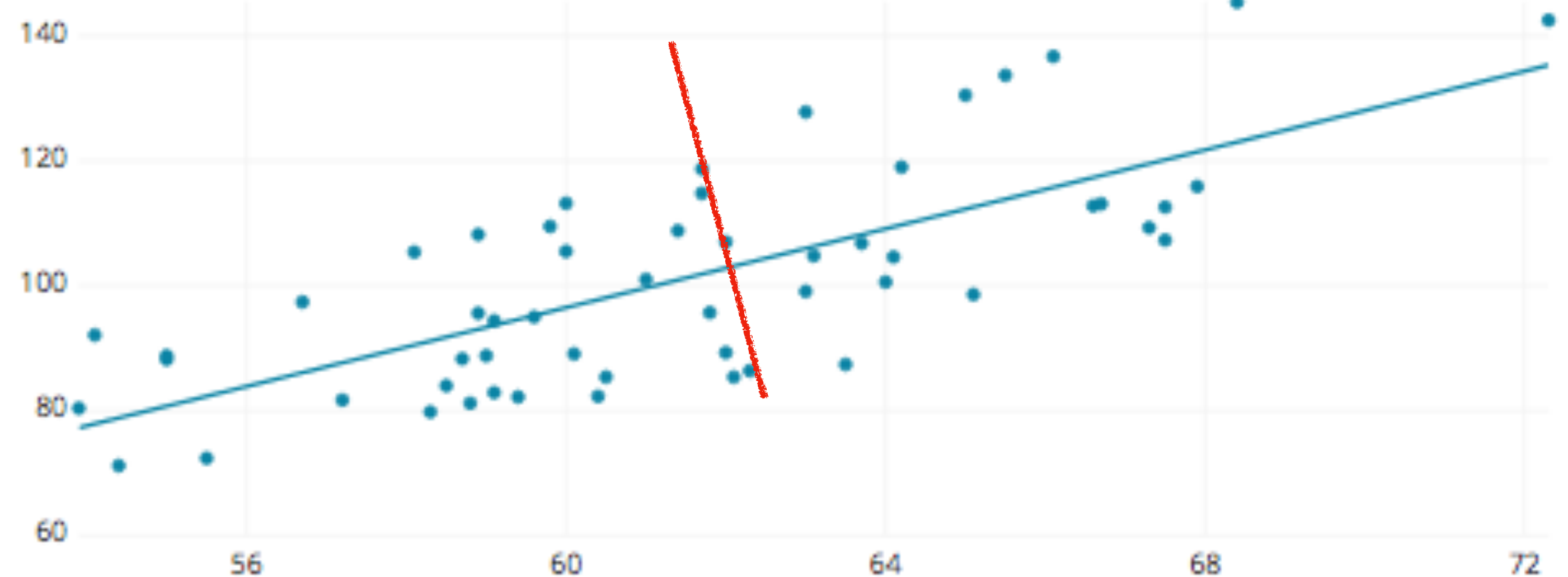
Intuición sobre PCA

- Ahora agreguemos una línea ortogonal.
- Empezamos a comprender lo de componentes



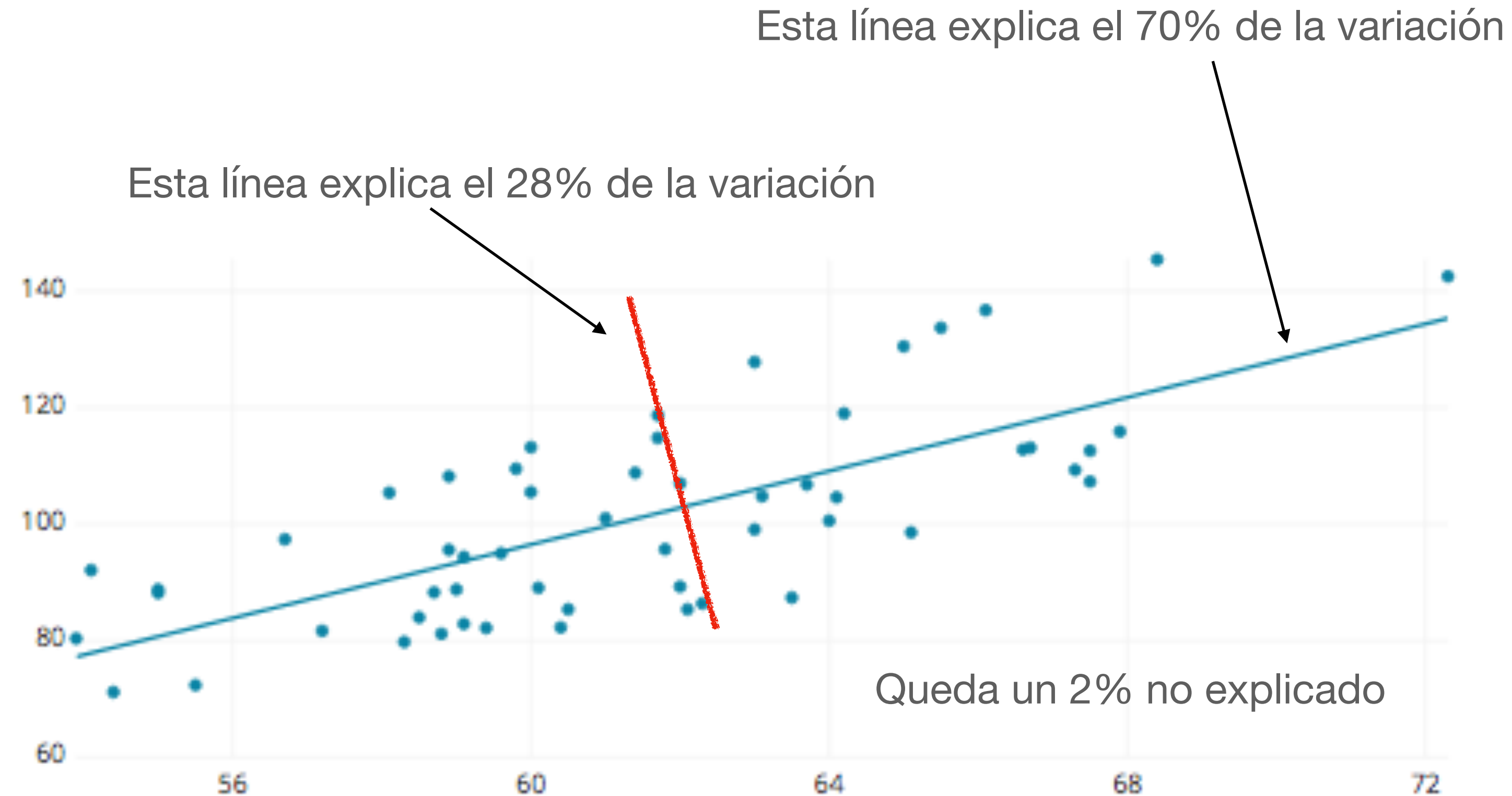
Intuición sobre PCA

Los componentes son una transformación lineal que selecciona un **nuevo** sistema de variables para el conjunto de datos tal que la varianza más grande del mismo cae sobre el primer eje.



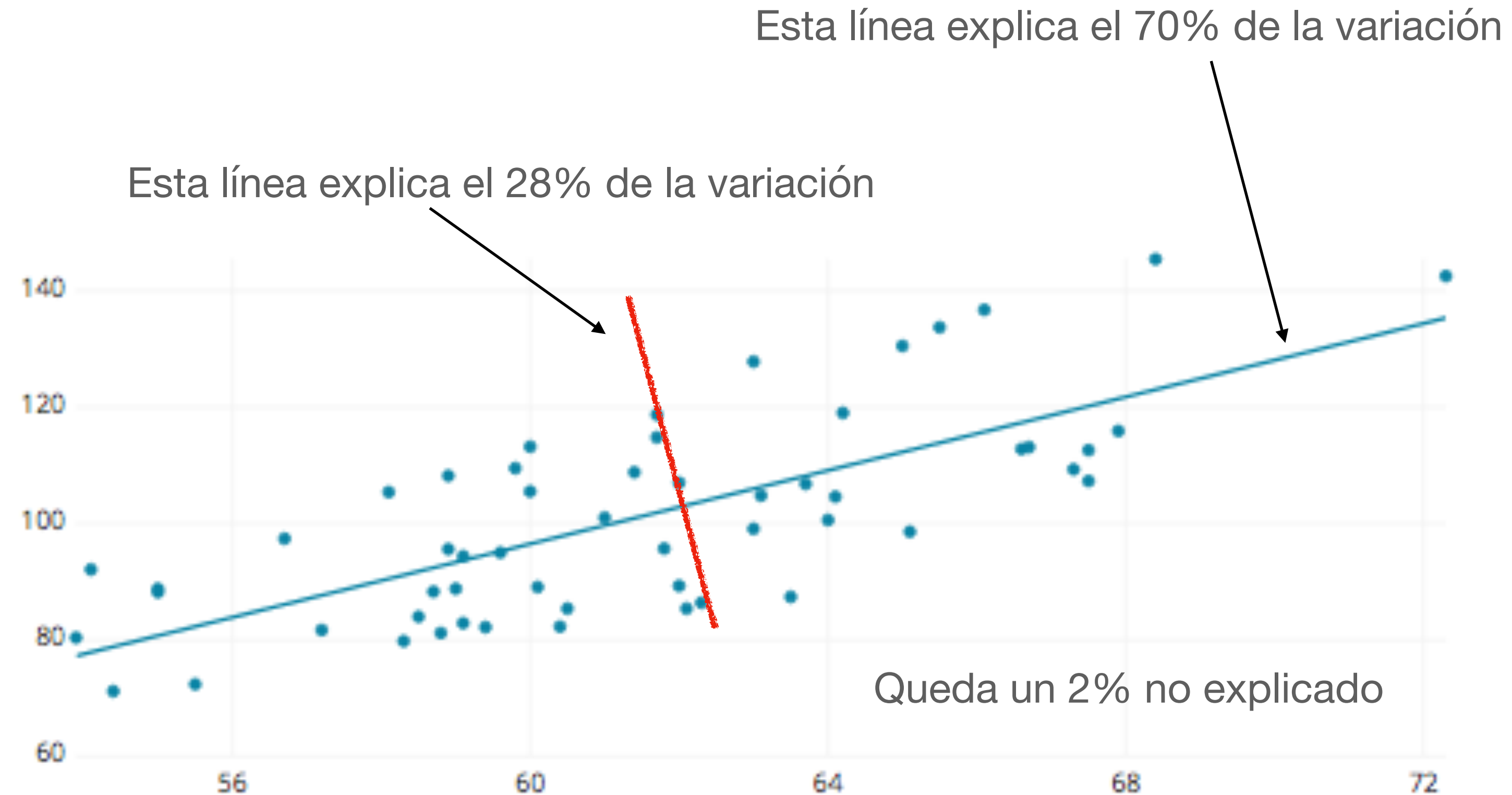
Intuición sobre PCA

- La segunda varianza mas grande cae sobre el segundo eje, y así nos vamos...
- Este proceso nos permite reducir el número de variables que se utilizan en un análisis.



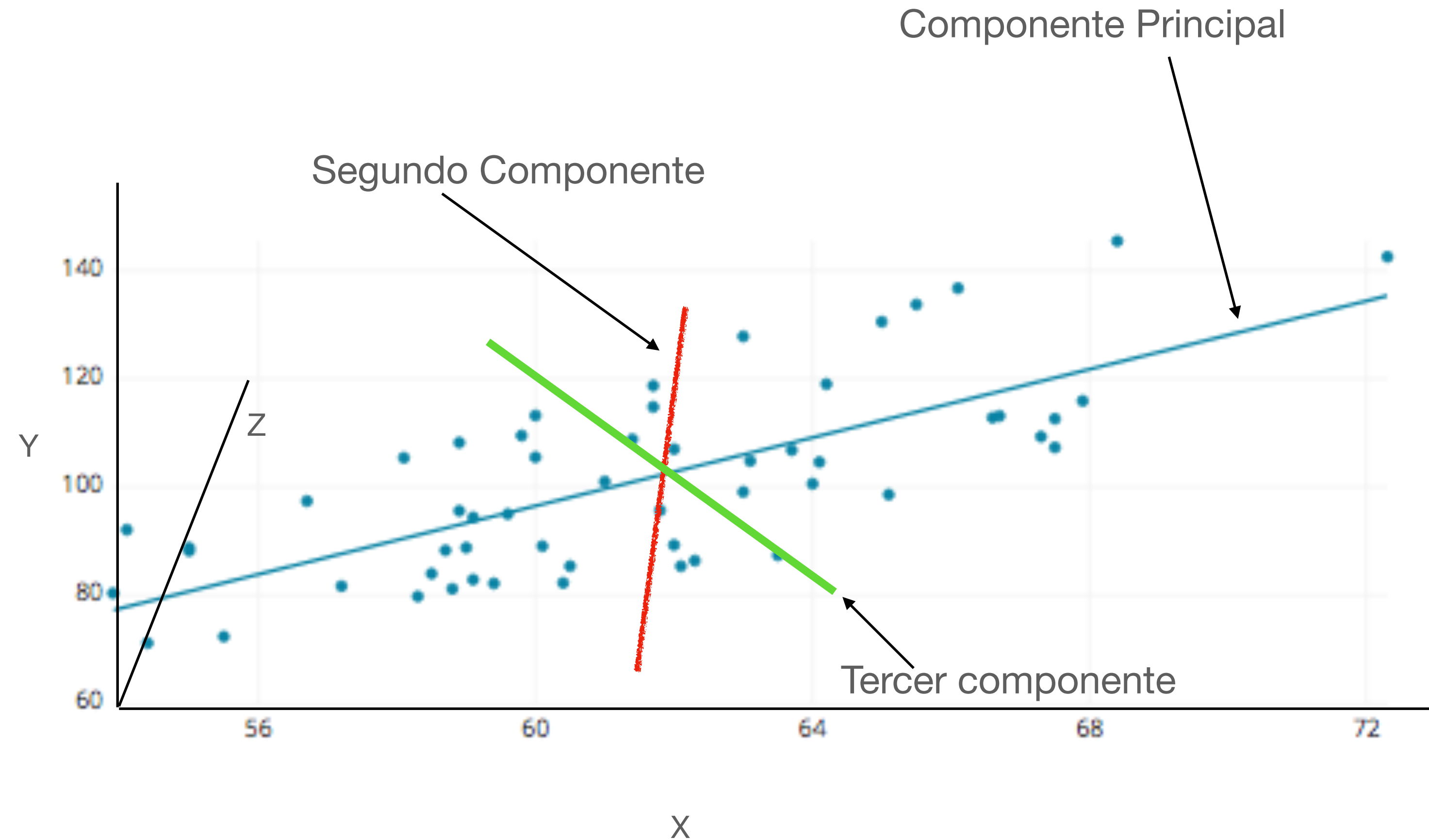
Intuición sobre PCA

Nótese que los componentes no están correlacionados ya que en este espacio están ortogonales entre sí.



Intuición sobre PCA

Podemos continuar este análisis hacia dimensiones más altas.



Intuición sobre PCA

- Si usamos esta técnica con un conjunto de datos que tiene un gran número de variables, podemos comprimir la cantidad de variación explicada a solo unos cuantos componentes.
- El reto mayor, y es el costo de hacer PCA, es la interpretación de los componentes.
 - Los componentes no se relacionan uno a uno con alguna de las variables
 - Los componentes son realmente combinaciones de las variables

Intuición sobre PCA

- Realizaremos un ejercicio de PCA con Python utilizando scikit learn.
- Para PCA, generalmente vamos a querer estandarizar nuestros datos a alguna escala, veremos como hacer esto también.
- El PCA es un algoritmo que usaremos generalmente para hacer análisis de datos, no es realmente un modelo por sí solo.