CENTER FOR URBAN SCIENCE+PROGRESS

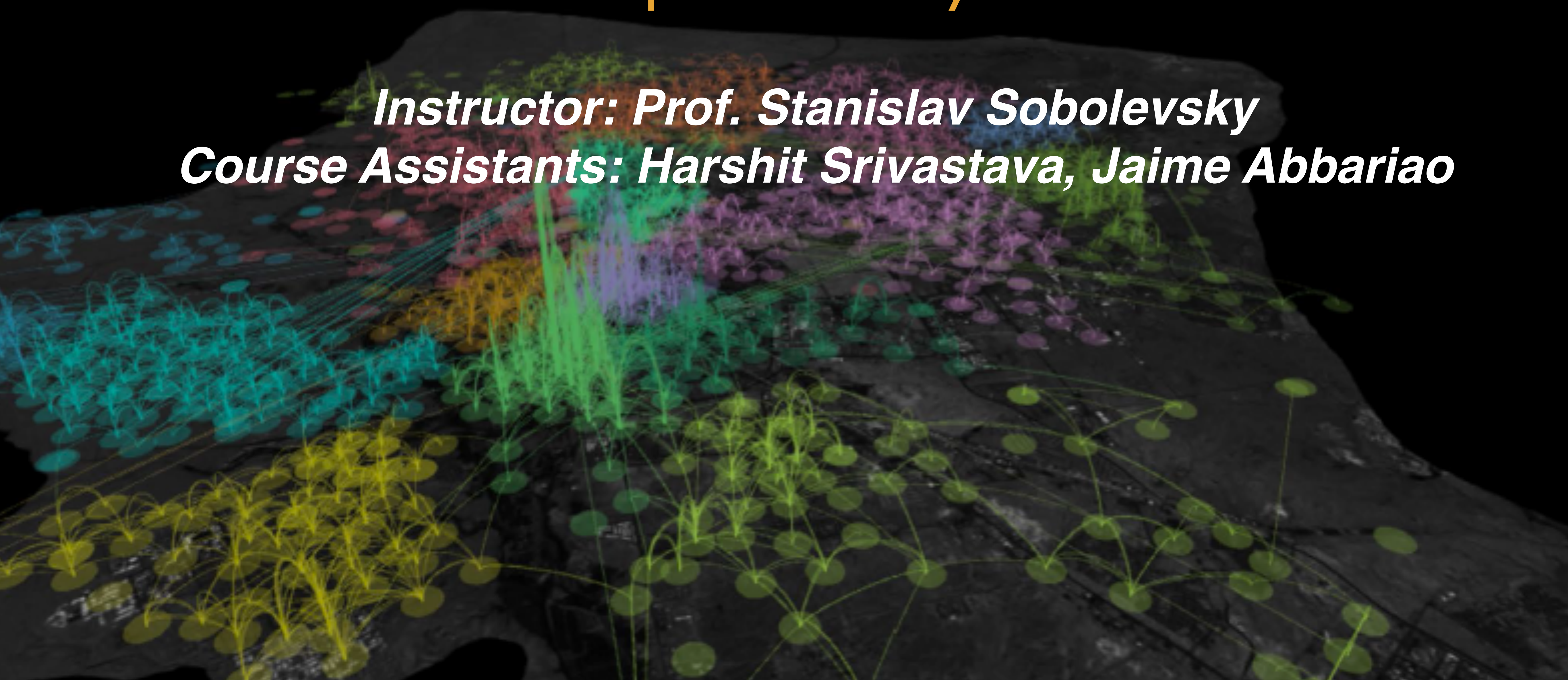
# APPLIED DATA SCIENCE

## 6004.002, Fall 2018

## Session 5: Dimensionality reduction. Principle component analysis

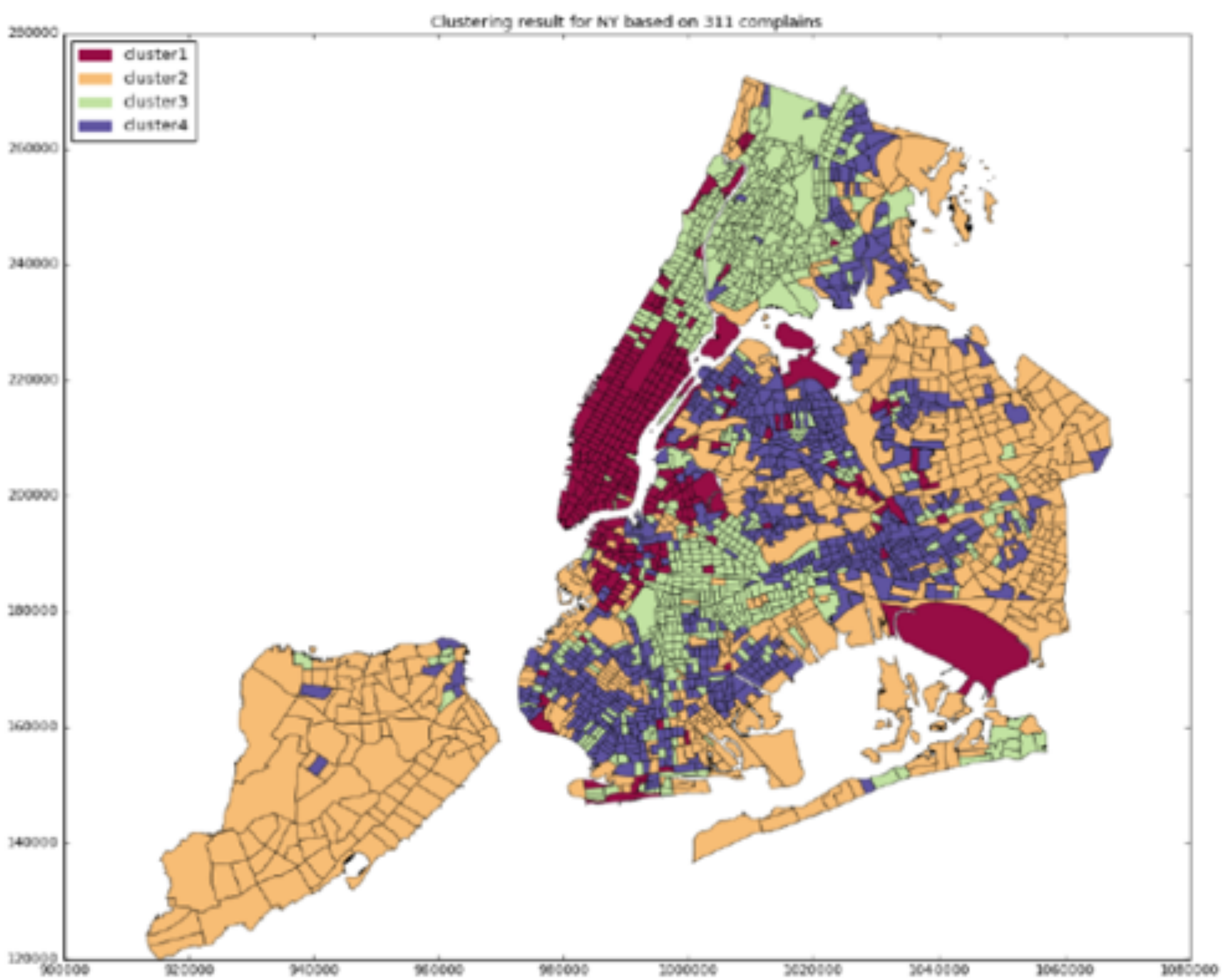

***Instructor: Prof. Stanislav Sobolevsky***

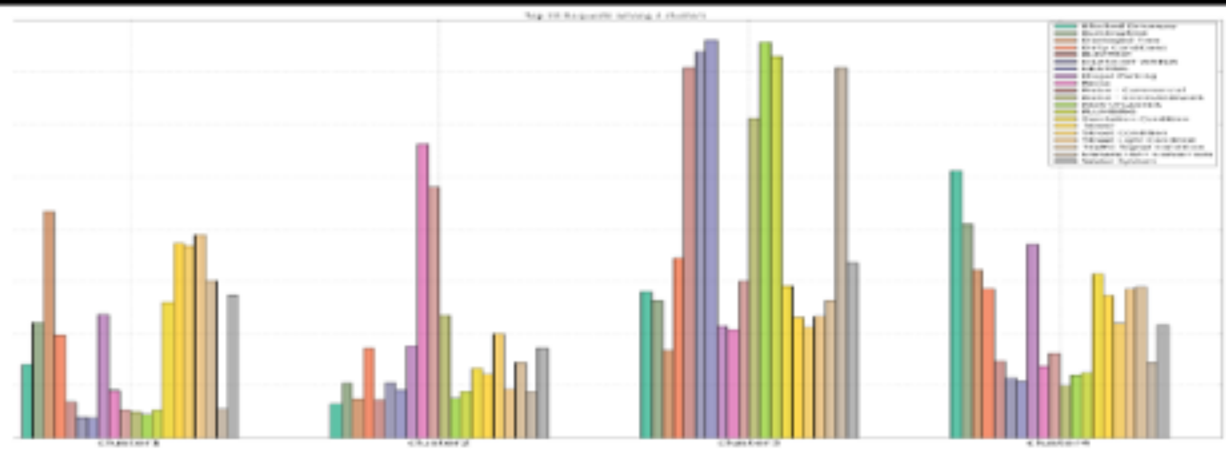
***Course Assistants: Harshit Srivastava, Jaime Abbariao***

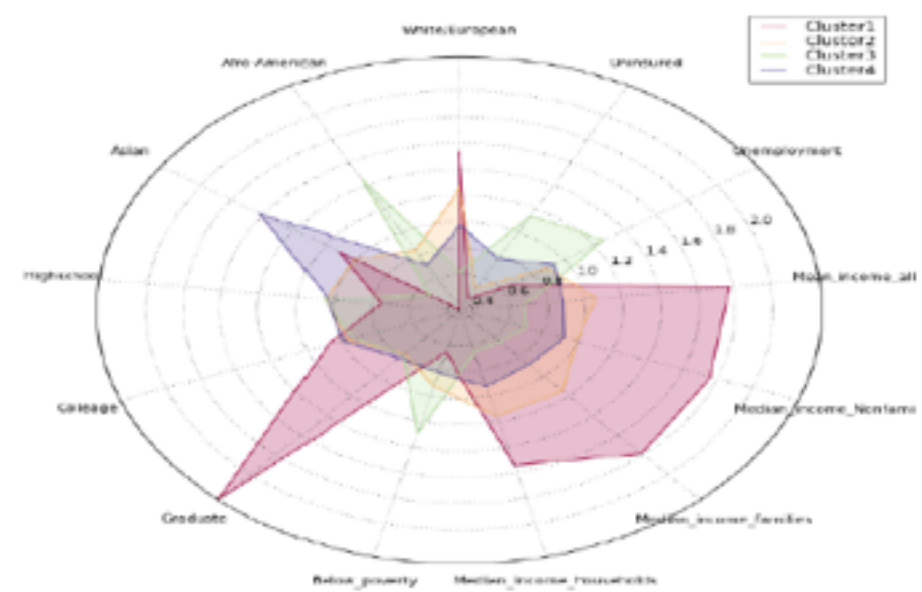



# Problem example

## Characterize urban neighborhoods with their 311 activity
## Model income, unemployment or average real estate prices






Wang L, Qian C, Kats P, Kontokosta C, **Sobolevsky, S.** (*corresp.) (2017) Structure of 311 service requests as a signature of urban location. PloS ONE. 12(10), e0186314.

# Issues with multi-dimensional data

- How to analyze/visualize it? $x = (x_1, x_2, x_3, ..., x_n)$

In case of regression: $y = f(x)$

- complexity
- irrelevant information
- multi-collinearity
- overfitting

# Skinnier data is often better

# Feature selection vs dimensionality reduction

$$y = f(x) \qquad\qquad x = (x_1, x_2, x_3, ..., x_n)$$

- feature selection reduces dimensionality of x by removing less relevant components

$$(x_1, x_2, x_3, x_4, x_5) \to (x_1, x_3, x_5)$$

- dimensionality reduction looks for more general mapping

$$(x_1, x_2, x_3, ..., x_n) \to (x'_1, x'_2, x'_3, ..., x'_m),\ m < n$$

$$y = f(x')$$

$$(x_1, x_2, x_3, x_4, x_5) \to x' = (x_1 + x_2 + x_3 + x_4 + x_5, x_1 x_2 x_3 x_4 x_5)$$

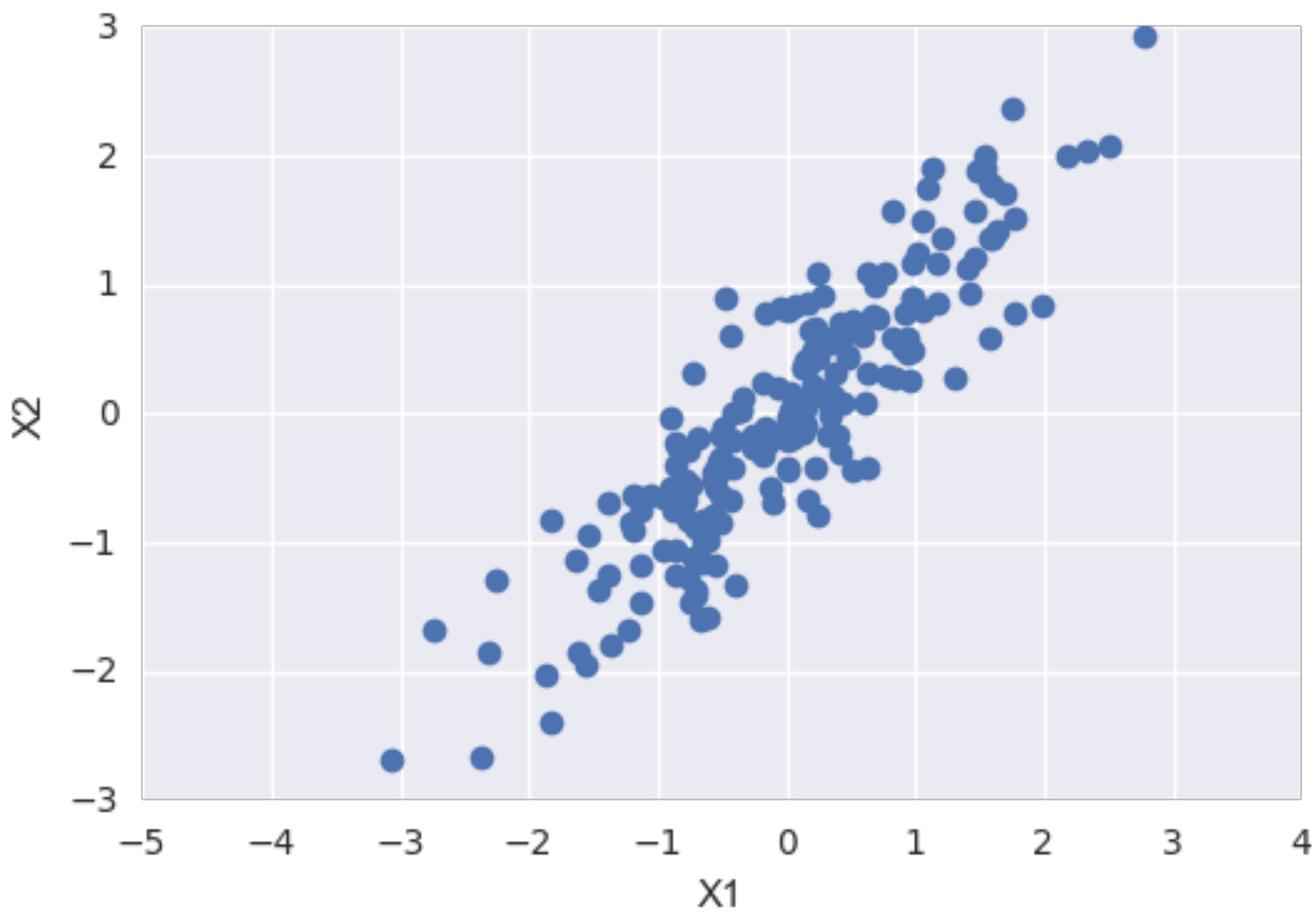
Pareto rule: 20% information often provides 80% of value

# Principal component analysis

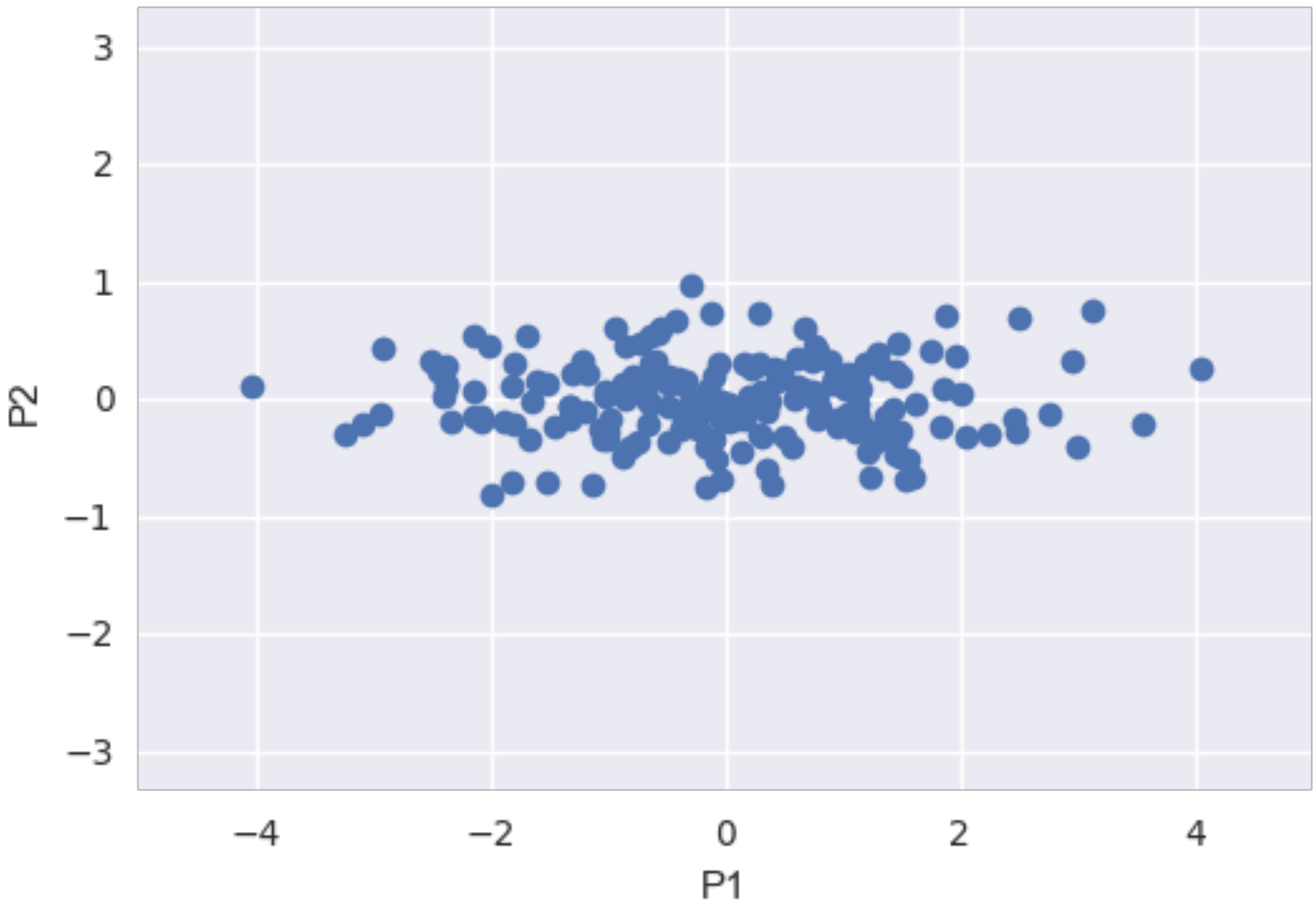
correlated features to uncorrelated

$$(x_1, x_2, x_3, ..., x_n) \rightarrow (u_1, u_2, u_3, ..., u_n)$$

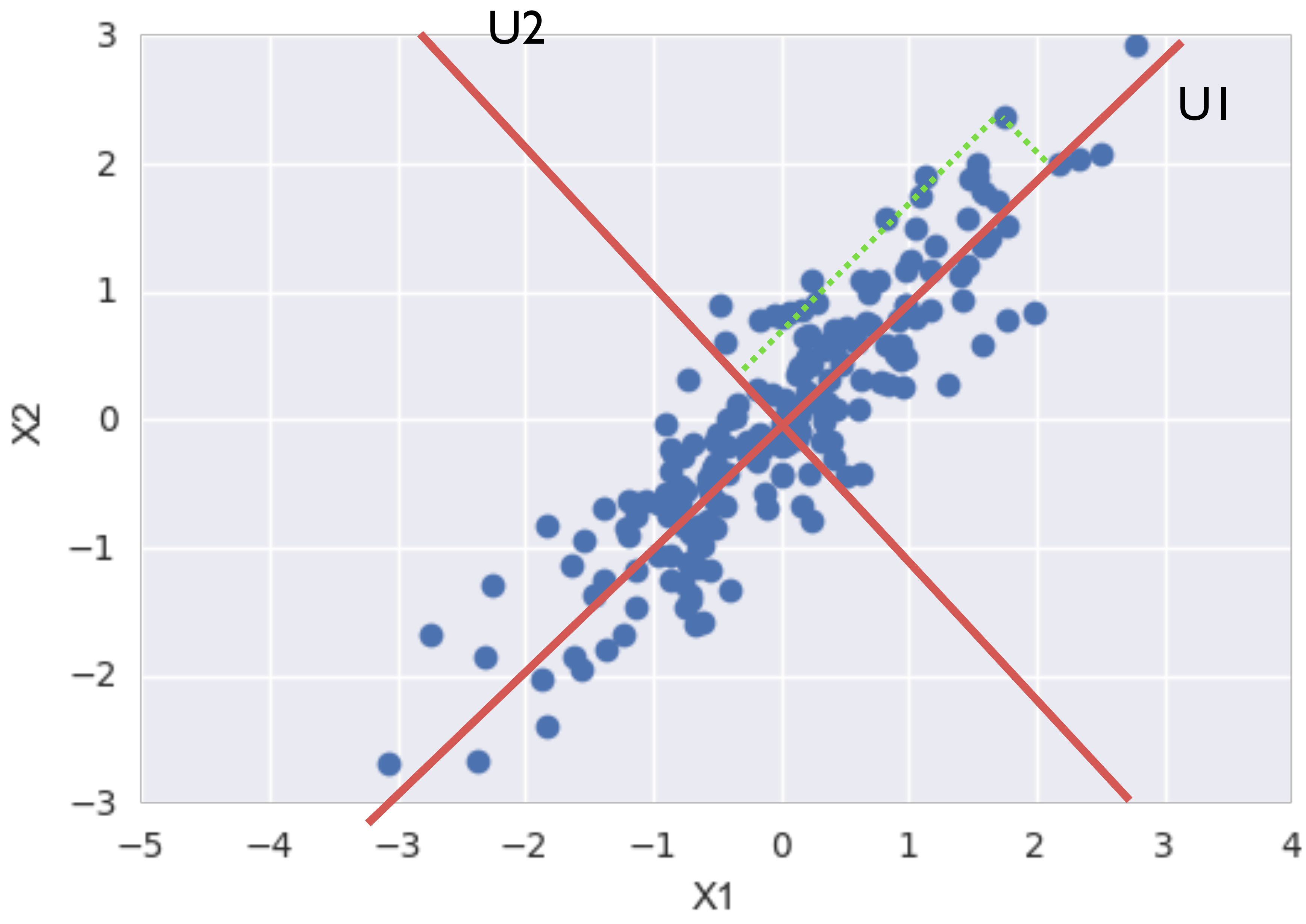
# Original data



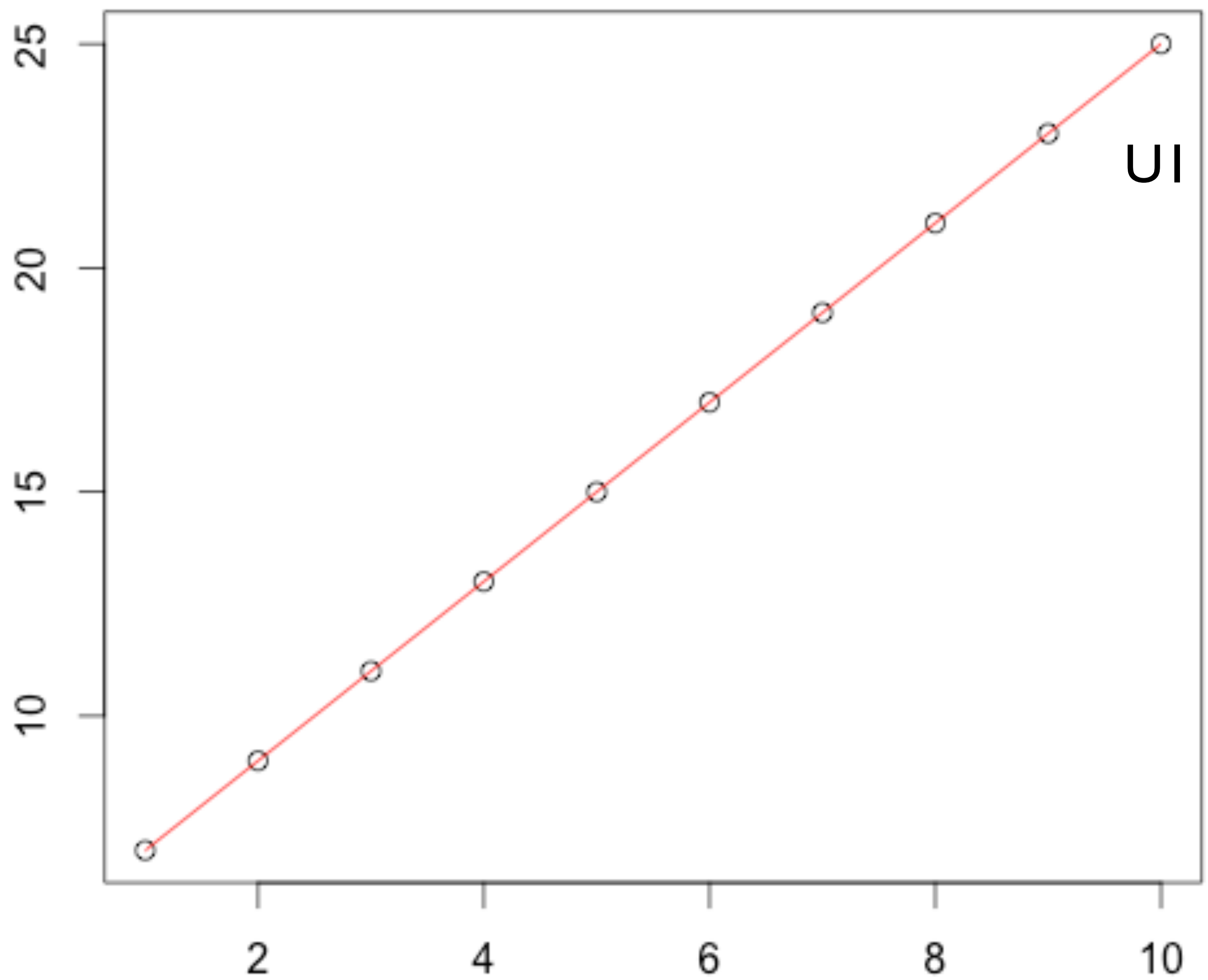
# Uncorrelated data (rotation)



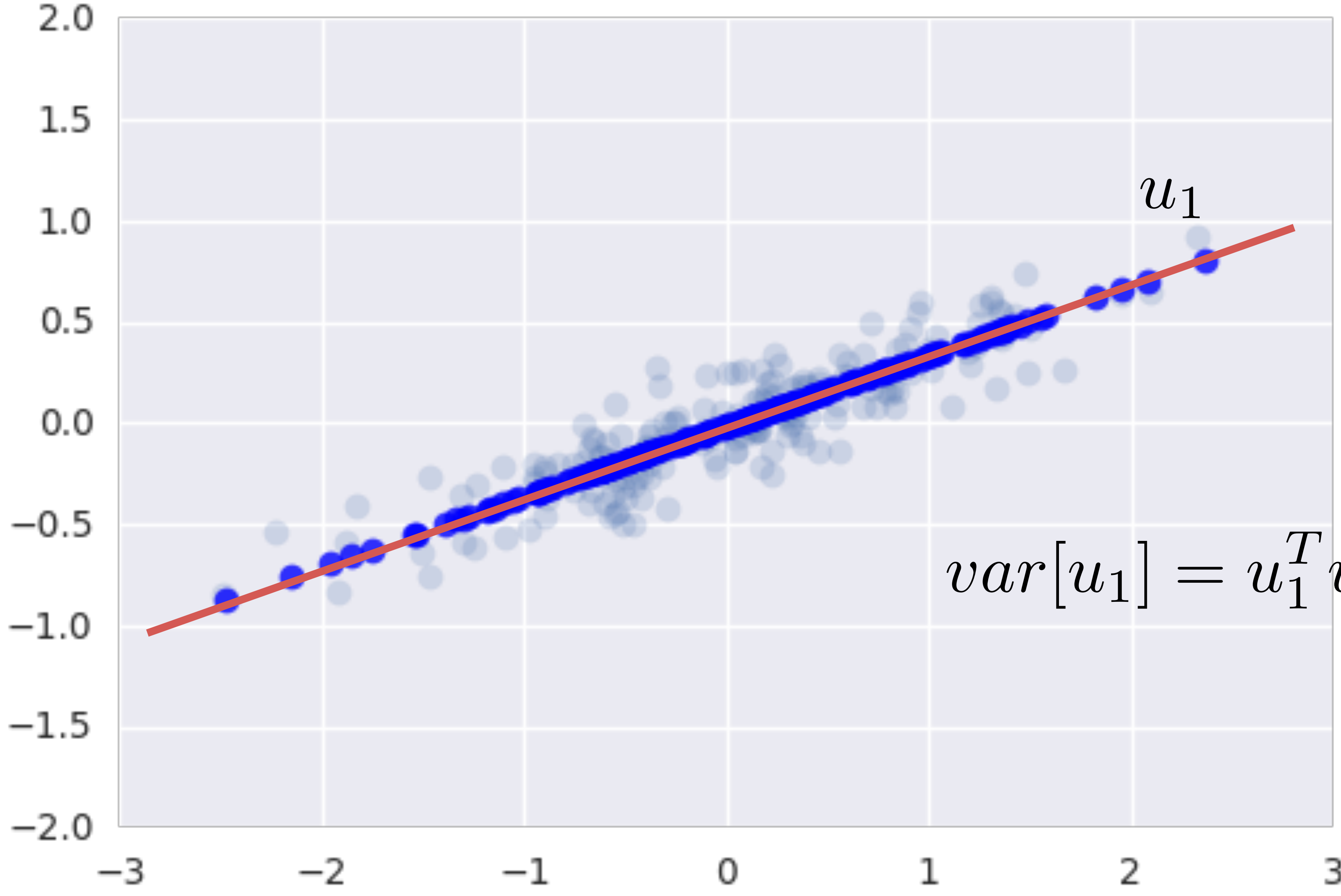
# Original data - new system of coordinates



# Same information with smaller number of parameters



# Almost the same information




$$var[u_1] = u_1^T u_1 \rightarrow max$$

# Principal components - maths

Given the standardized data $X = \{x_i^j, i = 1..n, j = 1..N\}$

Find uncorrelated latent factors U (or P)

$$u_j = x_1 v_j^1 + x_2 v_j^2 + ... + x_n v_j^n$$

Matrix for for the rotation transform

$$u_i = X v_i \qquad U = XV$$

V - n x p
U - N x p

Look for linear combinations of factors one-by-one

# Principal components - optimization objective

Looking for $u_j = x_1 v_j^1 + x_2 v_j^2 + \ldots + x_n v_j^n$

Start with $u_1 = X v_1$

Such that $var[u_1] = u_1^T u_1 \to max$

$$var[u_1] = u_1^T u_1 = (X v_1)^T (X v_1) = v_1^T X^T X v_1$$

Find $v_1 = argmax_{v_1 : v_1^T v_1 = 1} v_1^T X^T X v_1$

$$v_i = argmax_{v_i : v_i^T v_i = 1, v_i^T v_j = 0, j<i} v_i^T X^T X v_i$$

# Principal components - answer

So which vectors v maximize the quantity below?

$$v_i = argmax_{v_i : v_i^T v_i = 1, v_i^T v_j = 0, j<i} v_i^T X^T X v_i$$

Eigenvectors $\lambda_i v_i = X^T X v_i$ $\quad v_i^T v_i = 1 \quad v_i^T v_j = 0$

$$Var[u_i] = v_i^T X^T X v_i = \lambda_i v_i^T v_i = \lambda_i \quad \lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$$

Projection for the leading PC $v_1$

Is the leading eigenvector with the max eigenvalue

# Recall the concept of eigenvectors/eigenvalues

$$\lambda v = Av \qquad \lambda - eigenvalue, v - eigenvector$$

$$(\lambda I - A)v = 0 \qquad det(\lambda I - A) = 0$$

Find

$$\lambda_1, \lambda_2, ...\lambda_n$$
$$v_1, v_2, ...v_n$$

Define up to a scaling factor, can require unit length

$$v_i \rightarrow Cv_i \quad |v_i| = 1$$

When $\lambda_i \neq \lambda_j \quad \Rightarrow \quad v_i^T v_j = 0$

$$A^T = A$$

Proof $v_i^T A v_j = \lambda_j v_i^T v_j$

$$v_i^T A v_j = (Av_i)^T v_j = \lambda_i v_i^T v_j$$

# Principal components - proof

$$v_i = argmax_{v_i: v_i^T v_i = 1, v_i^T v_j = 0, j<i} v_i^T X^T X v_i$$

Consider eigenvectors:

$$\lambda_i v_i = X^T X v_i \qquad v_i^T v_i = 1 \qquad \lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$$

$$v_i^T X^T X v_i = \lambda_i v_i^T v_i = \lambda_i$$

$$w = e_1 v_1 + e_2 v_2 + \ldots + e_n v_n \qquad w^T w = e_1^2 + e_2^2 + \ldots + e_n^2 = 1$$

$$w^T X^T X w = \lambda_1 e_1^2 + \lambda_2 e_2^2 + \ldots + \lambda_n e_n^2 \rightarrow \max$$

$$w = v_1, e_1 = 1, e_2 = e_3 = \ldots = e_n = 0$$

# Principal components - select by variation

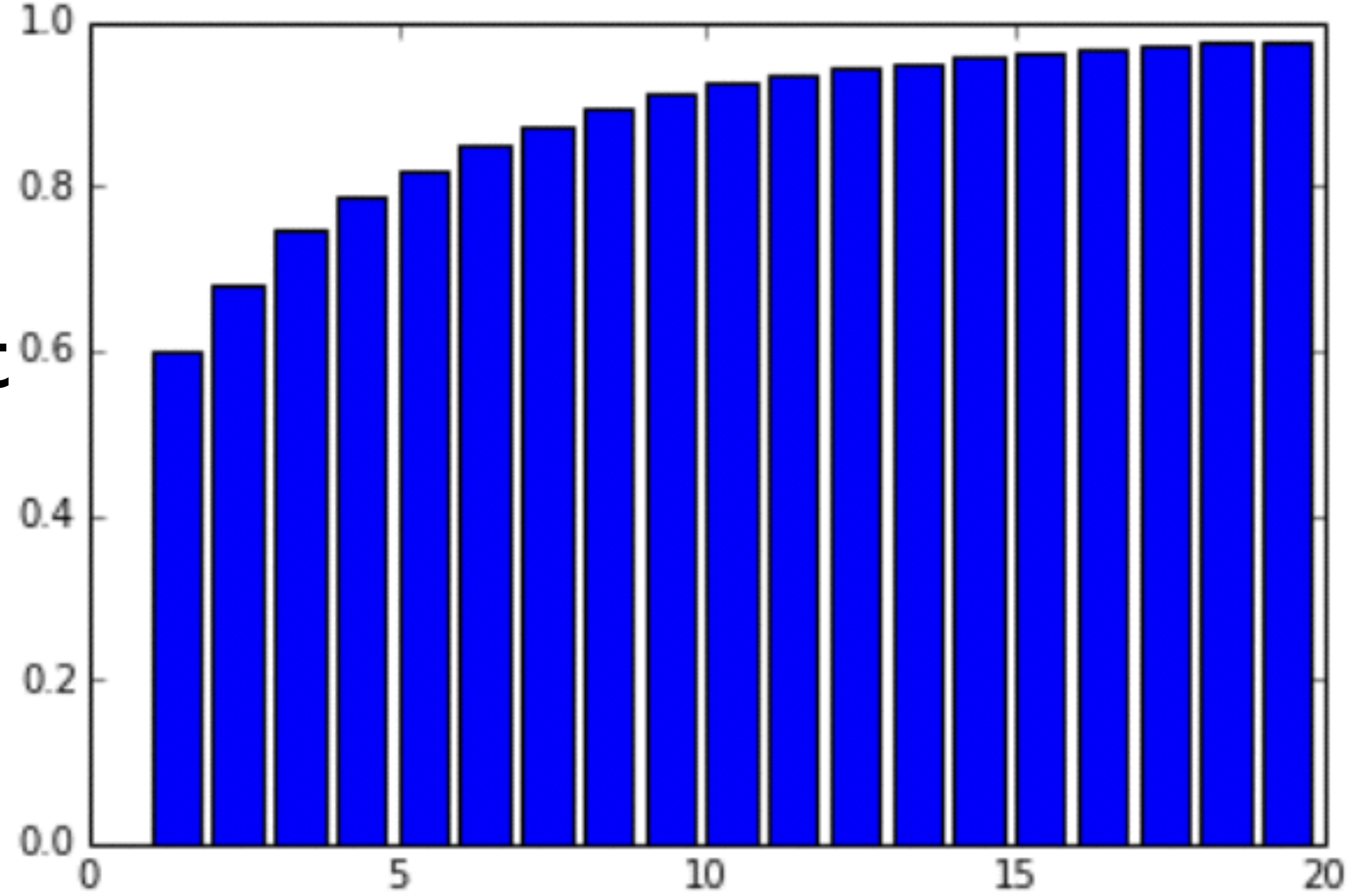
Information explained by a PC $u_i$

$$Var[u_i] = \lambda_i$$

% of total $\lambda_i / \sum_j \lambda_j$

Take first k, such that

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq \alpha$$




Information explained by leading PC's