Name	year	length
Carl Friedrich Gauss	1777	0
Camille Jordan	1838	12
Adrien-Marie Legendre	1752	0
Bernhard Riemann	1826	15
David Hilbert	1862	2
Henri Poincaré	1854	5
Emmy Noether	1882	0
Karl Weierstrass	1815	0
Eugenio Beltrami	1835	2
Hermann Schwarz	1843	20

We usually form the $n \times d$ matrix X whose ith row is X_i, with $1 \leq i \leq n$. Then the jth column is denoted by C_j ($1 \leq j \leq d$). It is sometimes called a *feature vector*, but this terminology is far from being universally accepted. In fact, many people in computer vision call the data points X_i feature vectors!

The purpose of *principal components analysis*, for short *PCA*, is to identify patterns in data and understand the *variance–covariance* structure of the data. This is useful for the following tasks:

1. Data reduction: Often much of the variability of the data can be accounted for by a smaller number of *principal components*.

2. Interpretation: PCA can show relationships that were not previously suspected.

Given a vector (a *sample* of measurements) $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, recall that the *mean* (or *average*) $\overline{x}$ of x is given by

$$\overline{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

We let $x - \overline{x}$ denote the *centered data point*

$$x - \overline{x} = (x_1 - \overline{x}, \ldots, x_n - \overline{x}).$$

In order to *measure the spread* of the x_i's around the mean, we define the *sample variance* (for short, *variance*) $\mathrm{var}(x)$ (or s^2) of the sample x by

$$\mathrm{var}(x) = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}.$$

Example 23.6. If $x = (1, 3, -1)$, $\overline{x} = \frac{1+3-1}{3} = 1$, $x - \overline{x} = (0, 2, -2)$, and $\mathrm{var}(x) = \frac{0^2+2^2+(-2)^2}{2} = 4$. If $y = (1, 2, 3)$, $\overline{y} = \frac{1+2+3}{3} = 2$, $y - \overline{y} = (-1, 0, 1)$, and $\mathrm{var}(y) = \frac{(-1)^2+0^2+1^2}{2} = 2$.