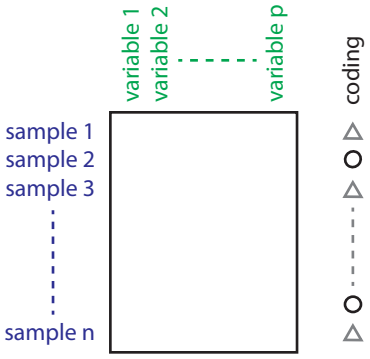
Visual Guide to PCA using Matlab

data



What combination of variables explain the differences between the ○ and the △ samples?

[coeff,score,~,~,explained] = pca(data)

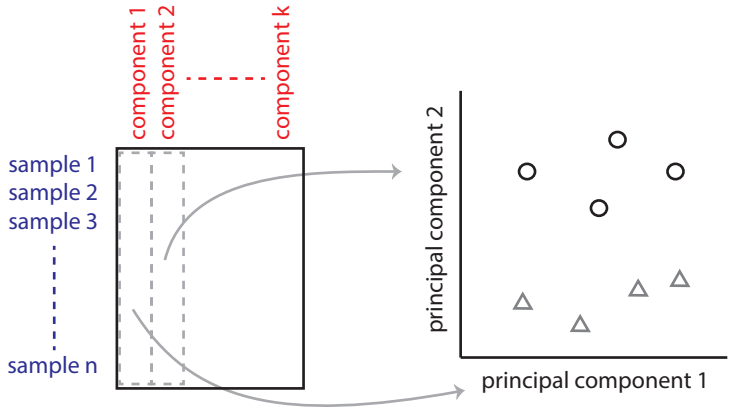
explained

component 1	35.26
component 2	23.02
component 3	8.94
component k	0.02

PCA returns up to k = min{n,p} principal components.

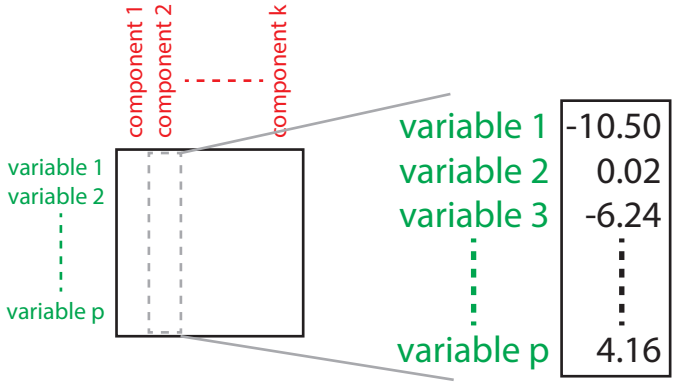
Principal component 1 explains 35.36% of the total variance. Principal component 2 explains another 23.02% of the total variance.

score



When plotted using principal components, the △ and ○ samples separate along principal component 2.

coeff



variable 1	-10.50
variable 2	0.02
variable 3	-6.24
variable p	4.16

The coefficients (or "loadings") connect the original variables and the principal components.

Variables are mean centered during PCA, so "low" samples are negative and "high" samples are positive.

Therefore, the ○ samples are low in variables 1 & 3 and high in variable p. The △ samples are low in variable p and high in variables 1 & 3.