1st Step: data pre-processing
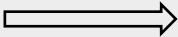
Raw dataset (input)
I_i: individual, $i = 1, ..., n$
V_j = categorical variable, $j = 1, ..., m$

	V_1	V_2	. . .	V_j
I_1				
I_2				
. . .				
I_i				

Multiple Correspondence Analysis

$\Longrightarrow$

Multiple Correspondence Analysis result
PC_j = Principal Component (continuous variable), $j = 1, ..., m$

	PC_1	PC_2	. . .	PC_j
I_1				
I_2				
. . .				
I_i				

Threshold
Inertia percentage
(at least 50%)

$\Longrightarrow$
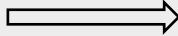
Dataset for Hierarchical clustering (output)
Selected PC, $k < j$
PC_k, $k = 1, ..., p$

	PC_1	PC_2	. . .	PC_k
I_1				
I_2				
. . .				
I_i				

2nd Step: Hierarchical Cluster Analysis

Dataset for Hierarchical clustering (input)
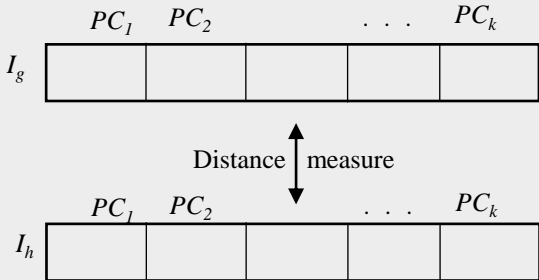Selected PC, $k < j$
PC_k, $k = 1, ..., p$

	PC_1	PC_2	. . .	PC_k
I_1				
I_2				
. . .				
I_i				

Hierarchical Cluster Analysis

$\Longrightarrow$
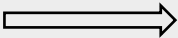
Starting agglomerative process:
distance between p-dimensional observations

(Euclidian distance)

	PC_1	PC_2	. . .	PC_k
I_g				

Distance measure $\updownarrow$

	PC_1	PC_2	. . .	PC_k
I_h				

Ward's algorithm
Agglomeration of clusters

$\Longrightarrow$
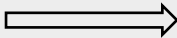
Hierarchical Cluster Analysis result
Hierarchical tree (output)