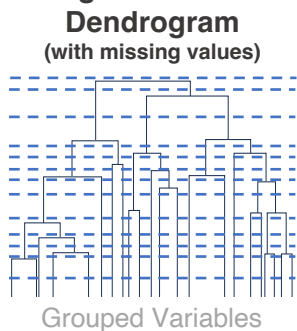


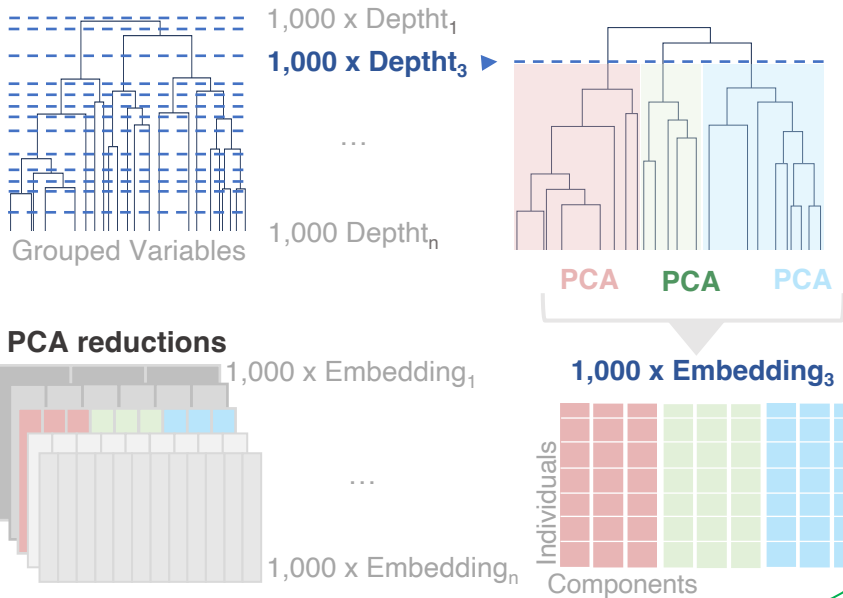
ClustALL – NA's

1. Data Complexity Reduction

A. Original Dataset



C.



B. 1,000x Imputation

1,000 complete Datasets

2. Stratification Process

A. For each embedding:

Distance metric + Clustering

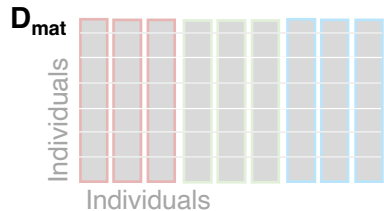
B. Evaluation of the optimal number of clusters

- Internal validation measurements
- Consensus across the imputed datasets for each *Depth*

Stratifications

Imputation + Embedding + Distance + Clustering

C.



Correlation Distance
+ H-cluster Clustering
+ Optimal cluster number

Final Stratifications

Embedding + Distance + Clustering

3. Consensus-based Stratifications