- Aggregation Combining two of more attributes (or objects) into a single attribute (or object) Purpose
 - Purpose
 - Data reduction
 - Reduce the number of attributes or objects
 - Change of scale
 - Cities aggregated into regions, states, countries, etc
 - More "stable" data
 - Aggregated data tends to have less variability

Sampling Sampling is the main technique employed for data selection. Prot is often and for both the preliminary investigation of the data and the final data analysis.

- Statisticians sample because obtaining the entire set of data of interest is too expensive or time consuming.
- Sampling is used in data mining because processing the entire set of data of interest is too expensive or time consuming.



Where *n* is the number of dimensions (attributes) and p_k and q_k are, respectively, the kth attributes (components) or data objects *p* and *q*.

• Standardization is necessary, if scales differ.