

archive: A physical or digital library that contains a collection of historical documents or records.

big data: Unusually large datasets that are collected digitally and, because of their variety and structure, may require sophisticated computational methods.

codebook: A guide for coding that consists of a list of the variables together with definitions, codes, and instructions for applying the codes.

comparative historical analysis: The development of causal explanations of social change by describing and comparing historical processes within and across cases.

content analysis: Systematic analysis of the symbolic content of communications in which the content is reduced to a set of coded variables or categories.

data archives: Repositories of survey, ethnographic, or qualitative interview data collected by various agencies and researchers that are accessible to the public.

demography: The study of the structure of and changes in human populations.

dictionary: In computerized content analysis, the set of words, phrases, or other word-based indicators (e.g., word length) that is the basis for a search of texts.

ecological fallacy: Erroneous use of data describing an aggregate unit (e.g., organizations) to draw inferences about the units of analysis that comprise the aggregate (e.g., individual members of organizations).

narrative comparison: A method of causal inference in which historical narratives of cases are analyzed to develop a general cross-case causal pattern and to validate it within each case.

process analysis: A within-case method of comparative historical analysis that examines possible intervening mechanisms that link an observed or theoretical association between events. Also called *process tracing*.

recording units: The units of analysis in content analysis, such as words, sentences, paragraphs, and whole articles. Also called *coding units*.

selective deposit: Systematic biases in the content of existing historical data due to actions such as selective destruction or editing of written records.

selective survival: Incompleteness of existing historical data due to the fact that some objects survive longer than others.

truth table: A table that presents all possible combinations of values, coded 0 (absent) or 1 (present), for a set of causal variables.

vital statistics: Data collected from the registration of “vital” life events, such as births, deaths, marriages, and divorces.