

Data

- Answering Who, What, When, Where, Why and How can provide a context for data values
 - If you can't answer who and what then you don't have data
 - Add the context of Who and What and organize the values into a data table
- The rows of a data table correspond to individual cases about whom (or about which- if they're not people) we record some characteristics
- Case- an individual about whom or which we have data.
- Respondents - individuals who answer a survey
- Subjects or Participants- people on whom we experiment
- Experimental Units- animals, plants, websites, and other inanimate subjects that are tested on
- Records- in a database rows are called records ex. Purchase records
- Variables- the characteristics about each individual case (What has been measured)
- Relational Database- two or more separate data tables linked together so that information can be merged across them.
- Categorical Variable- variable names that categorize and answer questions about how cases fall into those categories.
- Quantitative Variable- when a variable measured numerical values with units and the variable tells us about the quantity of what is measured.
- Units- tell us how each value has been measured
- There are exactly as many categories as individuals and only one individual in each category. This is called the Identifier variable.
- Nominal Variables- Categorical variables used only to name categories
- Ordinal Values- the numeric location of an item within an order
- Time Series- when you have the same variable measured at regular intervals over time. (popular in business)
- Ratio Level- Similar to interval, except there is a true zero, or starting point, and the ratios of data values have meaning.
- Cross-Sectional Data- where several variables are measured at the same time point
- Experiment- when you actively manipulate variables to see what happens.
- Observational Studies- Observations and measurements are made in a way that does not change the response or the variable being measured.

- Data Mining- the analysis of large transactional data sets for purposes other than those for which the data were originally collected.