

Chapter 2: Data

1. “Five W’s”: **Who, What, When, Where, and Why**: provide **context** for **data**.
 - a. **Data Table**: An arrangement of data where each row represents a **case** and each column represents a **variable**.
 - i. **Case**: An individual about whom or which we have data
 - ii. **Record**: Information about an individual in a database
 - b. **Relational Database**: Two or more separate data tables are linked together so that information can be merged across them; tables are called relations.
 - c. **Spreadsheet**: Layout designed for accounting that is often used to store and manage data tables

2. Types of **Variables**: the characteristics recorded about each individual or case
 - a. **Categorical Variable**: When a variable names categories and answers questions about how cases fall into these categories
 - i. **Nominal Variables**: Variables used only to name categories
 - ii. **Ordinal Values**: Values can be individually ordered
 - b. **Quantitative Variable**: When a variable has measured numerical values with units and the variable tells us about the quantity of what is measured
 - i. **Units**: Describe how each variable has been measured
 1. How much of something we have or how far apart the variables are
 - ii. **Time Series Data**: The same variable measured at regular intervals over time
 1. Months, quarters, years, etc
 - iii. ***Cross-Sectional Data**: Several variables are measured at the same time point
 - c. **Counts**: Used to summarize the frequency of each case of the CATEGORICAL variable occurring or they can be the values of a QUANTITATIVE variable whose units are a “number of something”
 - d. **Identifier Variable**: There are exactly as many categories as individuals and only one individual in each category; cannot analyze these numbers
 - i. *Example*: Student ID number: neither fully categorical nor quantitative.

3. Data Sources
 - a. **Surveys & Polls**: A study that asks questions of a sample drawn from some population in the hope of learning something about the entire population
 - i. **Respondents**: People who answer a survey
 - b. **Experiments**: Actively manipulate the variables, or factors, to see the results
 - i. **Subjects or Participants**: People who are experimented on
 - ii. **Experimental Unit**: Animals, plants, websites, and other inanimate subjects
 - c. **Observational Studies**: A study based on data in which no manipulation of factors has been employed
 - d. **Data Mining**: The analysis of large transactional data sets for purposes other than those for which the data were originally collected.