

# CH 1: Data and Statistics

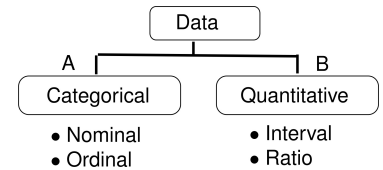
## 1. Key Definitions:

- a. Elements: The entities on which data are collected.
- b. Population (universe): The set of all elements of interest in a particular study. ( size  $N$ )
- c. Sample: A subset of the population selected for analysis. (size  $n$ )  
EX 1: ID the population and the sample for the Gallup Poll (estimation of the percentage of popular vote for each Candidate based on interviews of 1500 adults)
- d. Census: A survey to collect data on the entire population.
- e. Sample survey: A survey to collect data on a sample.
- f. Variable: A characteristic of interest for the elements.

## 2. Scales of Measurement

- a. Nominal: classification data (e.g. m/f); no ordering (e.g. it makes no sense to state that  $m > f$ ; arbitrary labels (e.g., m/f, 0/1, etc).
- b. Ordinal: ordered but differences between values are not important (e.g., political parties on left to right spectrum given labels 1,1,2; restaurant ratings)
- c. Interval: ordered, constant scale, but no natural zero; differences make sense, but ratios do not (e.g., temperature:  $30^{\circ} - 20^{\circ} = 20^{\circ} - 10^{\circ}$ , but  $\frac{20^{\circ}}{10^{\circ}}$  is not twice as hot).
- d. Ratio: ordered, constant scale, natural zero (e.g., height, weight, age, length)

## 3. Data Structure



A Data classified in categories. e.g. Gender, Hair Color, Bond Rating.

B Data measured on numerical scale. e.g. Age, temperature.

EX2 Table 1.6 shows the fuel efficiency ratings for 10 cars.

Car	Class	Cylinders	City MPG	Highway MPG	Fuel Type
Audi A8	Large	12	13	19	Premium
BMW 328Xi	Compact	6	17	25	Premium
Cadillac CTS	Midsize	6	16	25	Regular
Chevrolet Malibu	Midsize	6	17	26	Regular
Chrysler 300	Large	8	13	18	Premium
Ford Focus	Compact	4	24	33	Regular
Hyundai Elantra	Midsize	4	25	33	Regular
Pontiac G6	Compact	6	15	22	Regular
Toyota Camry	Midsize	4	21	31	Regular
Volkswagen Jetta	Compact	5	21	29	Regular

- a How many elements are in this data set?
- b How many variables are in this data set?
- c ID the data structure of the variables.
- d ID the measurement scale of the variables.