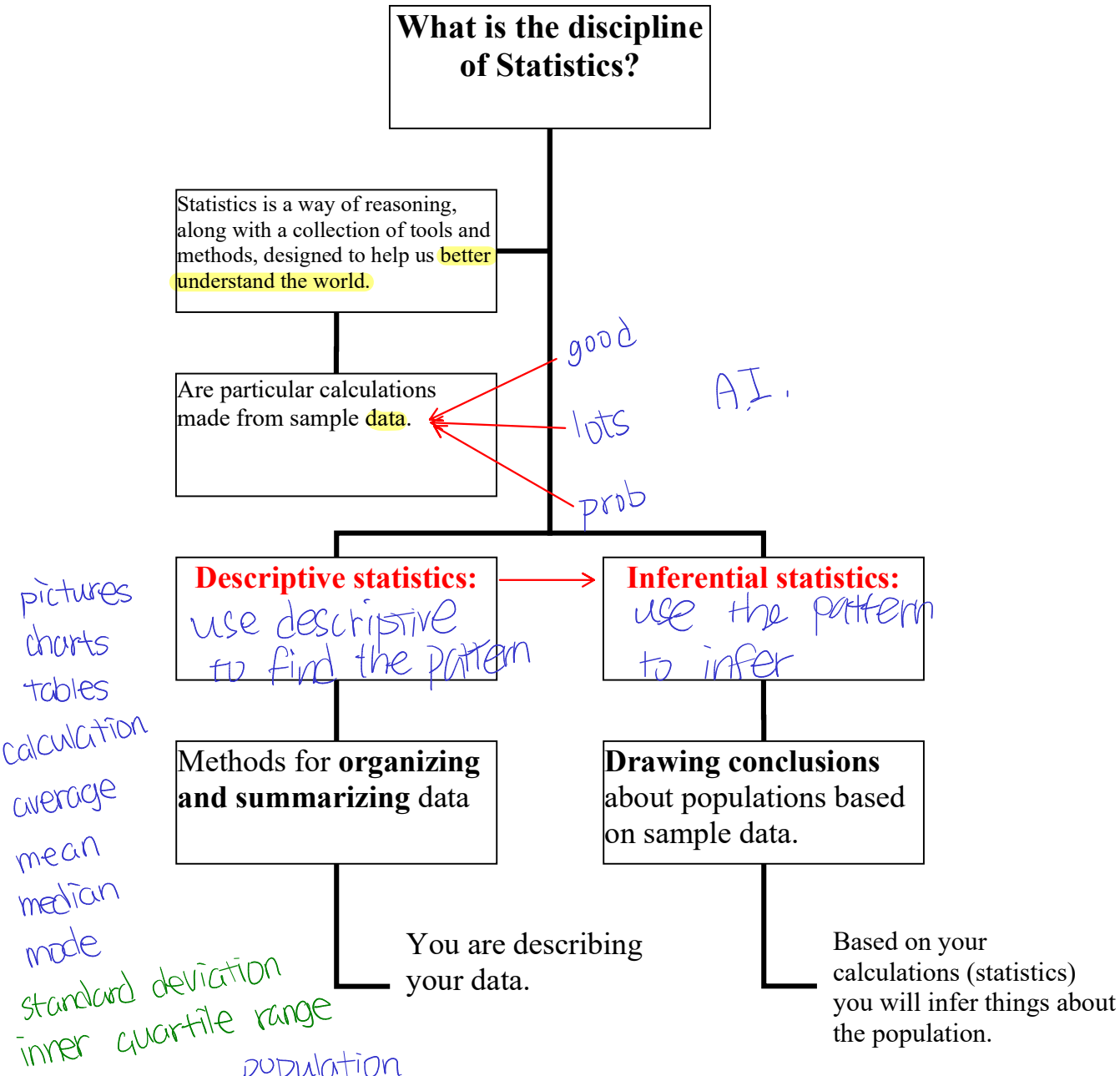
Lecture Notes

Chapters 1-2

Turning Data Into Information



- **Sample Size**

  population or objects universal letter : n

  How many people are in your sample?

- **Data**

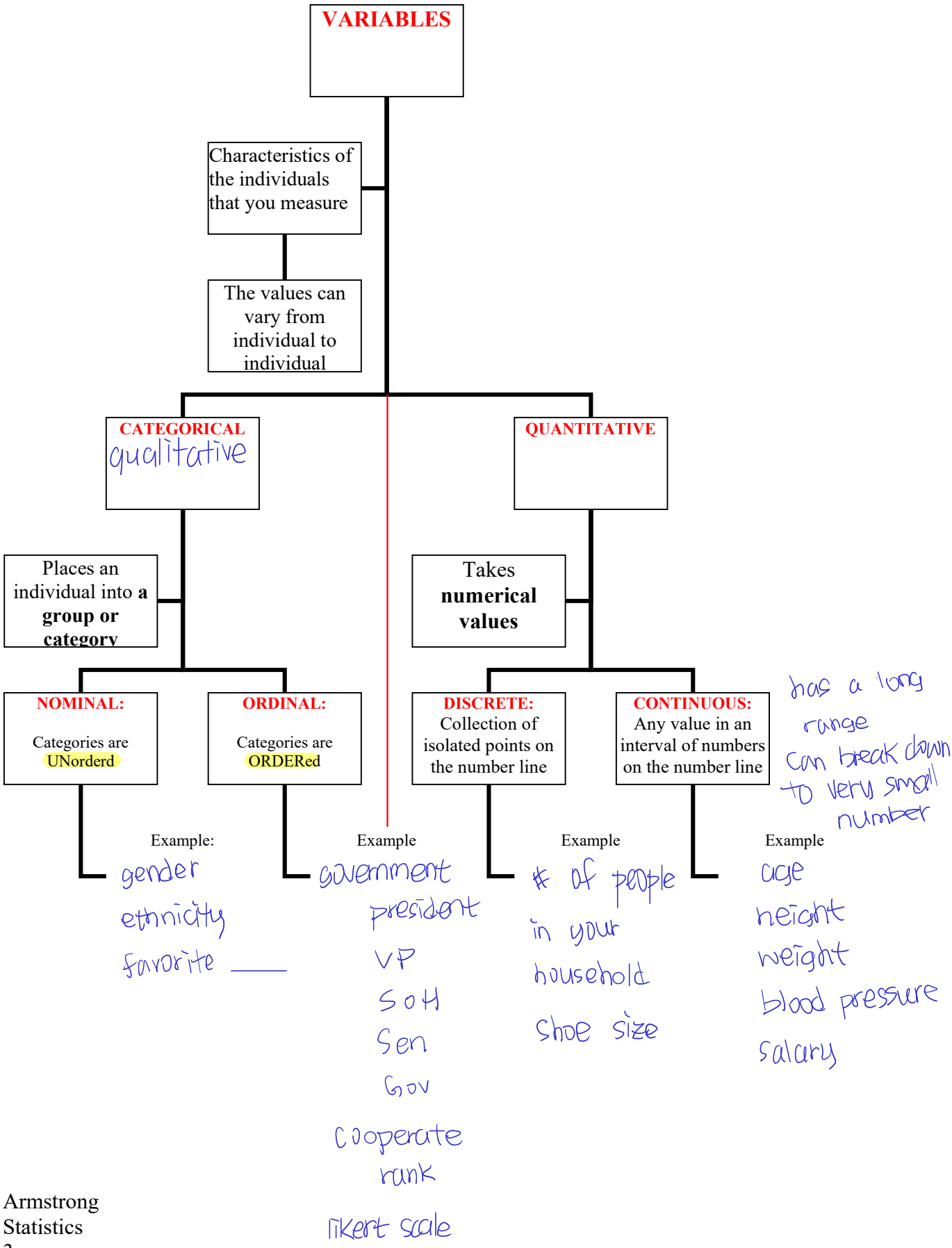
  Are values you collect (this could be height, color of walls, count of cars, etc)

  non - human

- **Experimental Units or Subjects** ← human

  Could be individuals or subjects under study

  These are the **people or objects** we gather information about.


VARIABLES
Characteristics of the individuals that you measure
The values can vary from individual to individual
CATEGORICAL
qualitative
QUANTITATIVE
Places an individual into a group or category
Takes numerical values
NOMINAL:
Categories are UNorderd
ORDINAL:
Categories are ORDERed
DISCRETE:
Collection of isolated points on the number line
CONTINUOUS:
Any value in an interval of numbers on the number line
has a long range
Can break down to very small number
Example:
gender
ethnicity
favorite ____
Example
government
president
VP
SoH
Sen
Gov
Cooperate
rank
likert scale
Example
# of people in your household
shoe size
Example
age
height
weight
blood pressure
salary

POPULATION:

The entire collection of persons, things or objects you wish to study.

POPULATION PARAMETER:

A number or calculation that describes or summarizes a population.

mean , std dev , proportion

μ θ p

mu sigma

SAMPLE:

A subset of the population. The sample should be representative of the entire population.

SAMPLE STATISTIC:

A number or calculation that describes or summarizes a sample.

mean , std dev , proportion

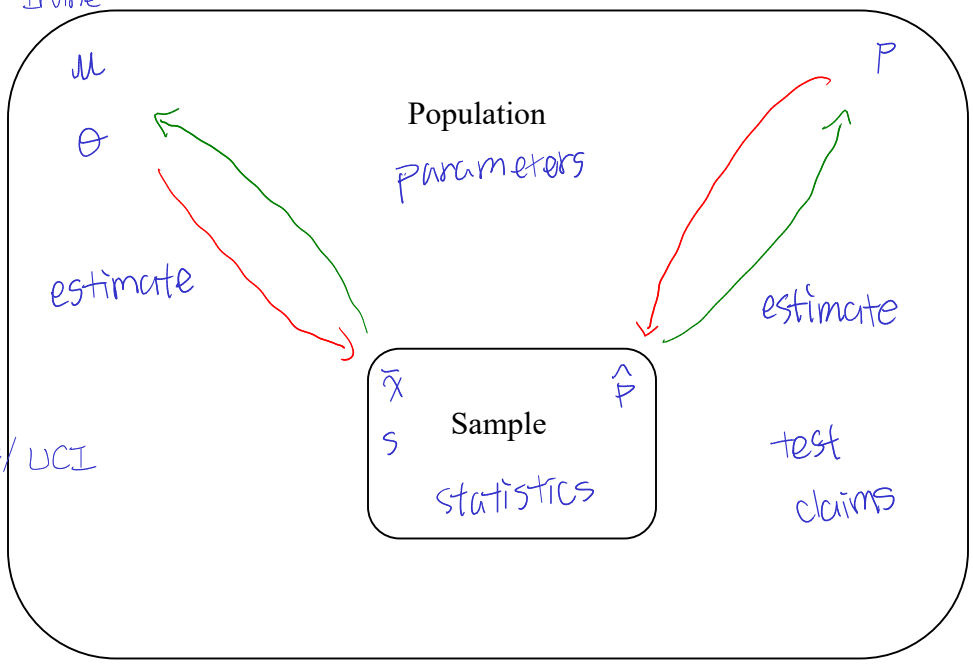
$\bar{x}$ s $\hat{p}$

x-bar p-hat

world / USA / Cali / Irvine

bad examples

USA / cali / Irvine / UCI



sample should mirror the population as much as possible

Sampling Variability: Each sample will select different people, and therefore, different values for the measured variables (no two samples will be identical)

The larger the sample size, the closer / better the $\bar{x}_1$ and $\bar{x}_2$ will be estimated of the true population

$\bar{x}_1 \approx \bar{x}_2$

n_1 n_2 ↑

Count / frequency
relative frequency

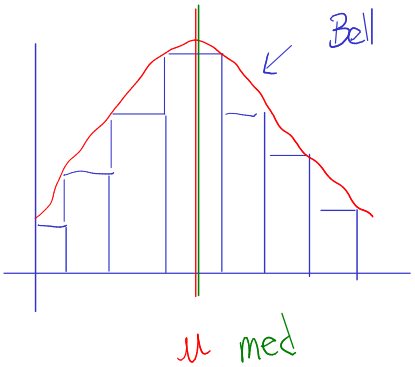
A DISTRIBUTION TELLS US WHAT VALUES A VARIABLE CAN TAKE AND HOW OFTEN IT TAKES THESE VALUES.

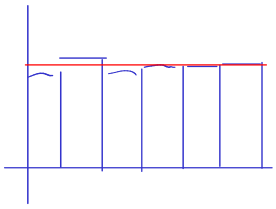
Histograms, box plots, bar charts, pie charts all help us understand what the distribution is of data is.

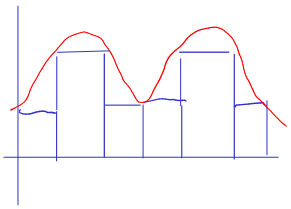
For quantitative data:
If the distribution is symmetric:
Center: Mean
Spread: Standard Deviation
Best Picture: Histogram

$$\theta = \sqrt{\frac{\Sigma(x_i - u)^2}{N}}$$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$







If the distribution is skewed:
Center: Median
Spread: IQR
Best Picture: Box Plot