

## **Chapter 1**

# **Introduction to Statistics**

## § 1.1

# An Overview of Statistics

# Data and Statistics

**Data** consists of information coming from observations, counts, measurements, or responses.

**Statistics** is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.

A **population** is the collection of *all* outcomes, responses, measurement, or counts that are of interest.

A **sample** is a subset of a population.

# Populations & Samples

## Example:

In a recent survey, 250 college students at Union College were asked if they smoked cigarettes regularly. 35 of the students said yes. Identify the population and the sample.

Responses of all students at  
Union College (**population**)

Responses of students  
in survey (**sample**)

# Parameters & Statistics

A **parameter** is a numerical description of a *population* characteristic. Parameters are denoted by Greek letters

A **statistic** is a numerical description of a *sample* characteristic. Statistics are denoted by Latin letters

Parameter → Population

Statistic → Sample

# Parameters & Statistics

## Example:

Decide whether the numerical value describes a population parameter or a sample statistic.

- a.) A recent survey of a sample of 450 college students reported that the average weekly income for students is \$325.

Because the average of \$325 is based on a sample, this is a sample statistic.

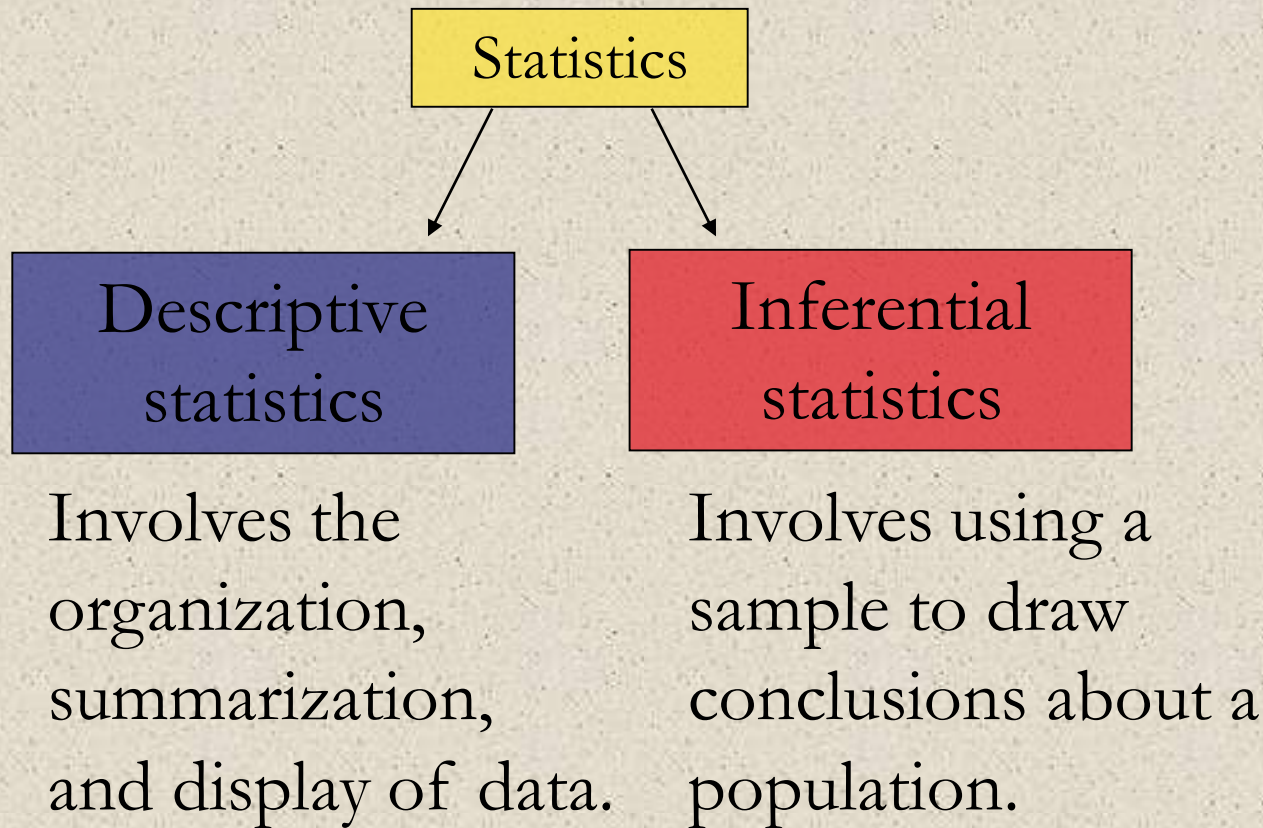
- b.) The average weekly income for all students is \$405.

Because the average of \$405 is based on a population, this is a population parameter.



# Branches of Statistics

The study of statistics has two major branches: **descriptive statistics** and **inferential statistics**.



# Descriptive and Inferential Statistics

## Example:

In a recent study, volunteers who had less than 6 hours of sleep were four times more likely to answer incorrectly on a science test than were participants who had at least 8 hours of sleep. Decide which part is the descriptive statistic and what conclusion might be drawn using inferential statistics.

The statement “four times more likely to answer incorrectly” is a descriptive statistic. An inference drawn from the sample is that all individuals sleeping less than 6 hours are more likely to answer science question incorrectly than individuals who sleep at least 8 hours.

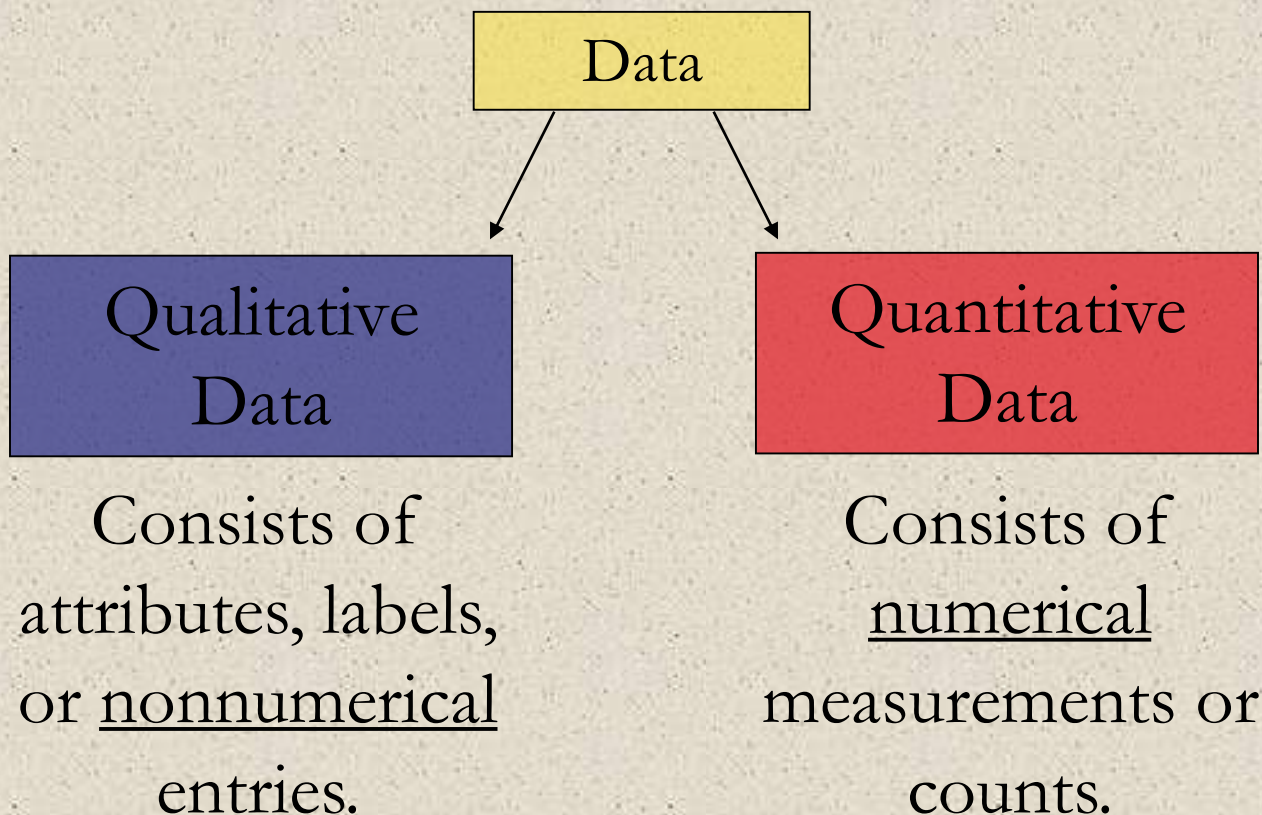


§ 1.2

# Data Classification

# Types of Data

Data sets can consist of two types of data: **qualitative data** and **quantitative data**.



# Qualitative and Quantitative Data

## Example:

The grade point averages of five students are listed in the table. Which data are qualitative data and which are quantitative data?

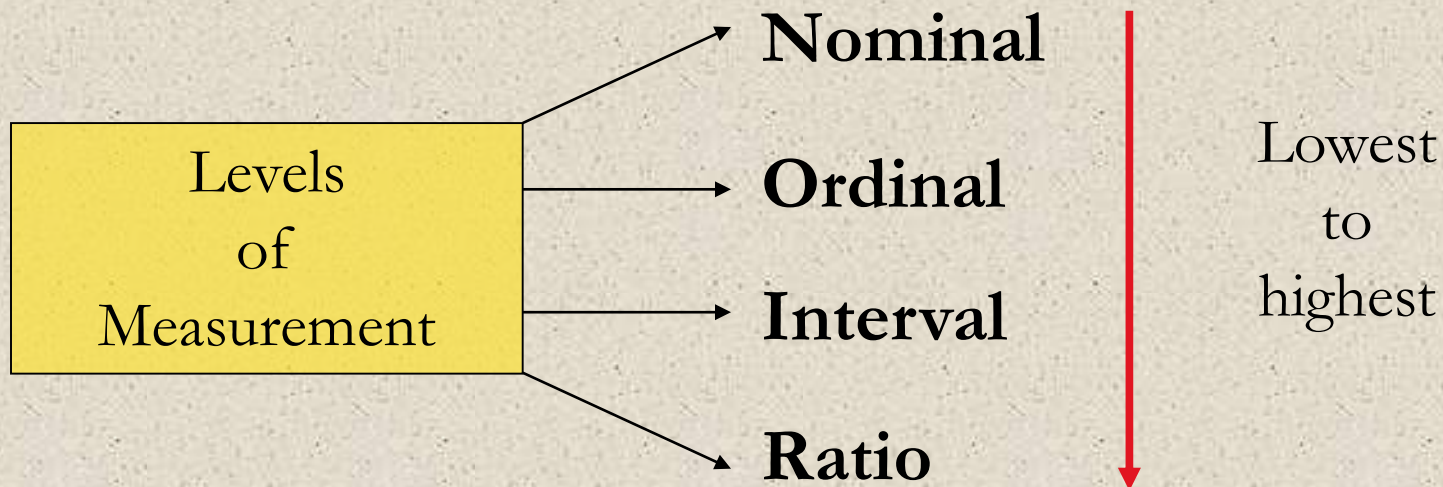
Student	GPA
Sally	3.22
Bob	3.98
Cindy	2.75
Mark	2.24
Kathy	3.84

Qualitative data ←

→ Quantitative data

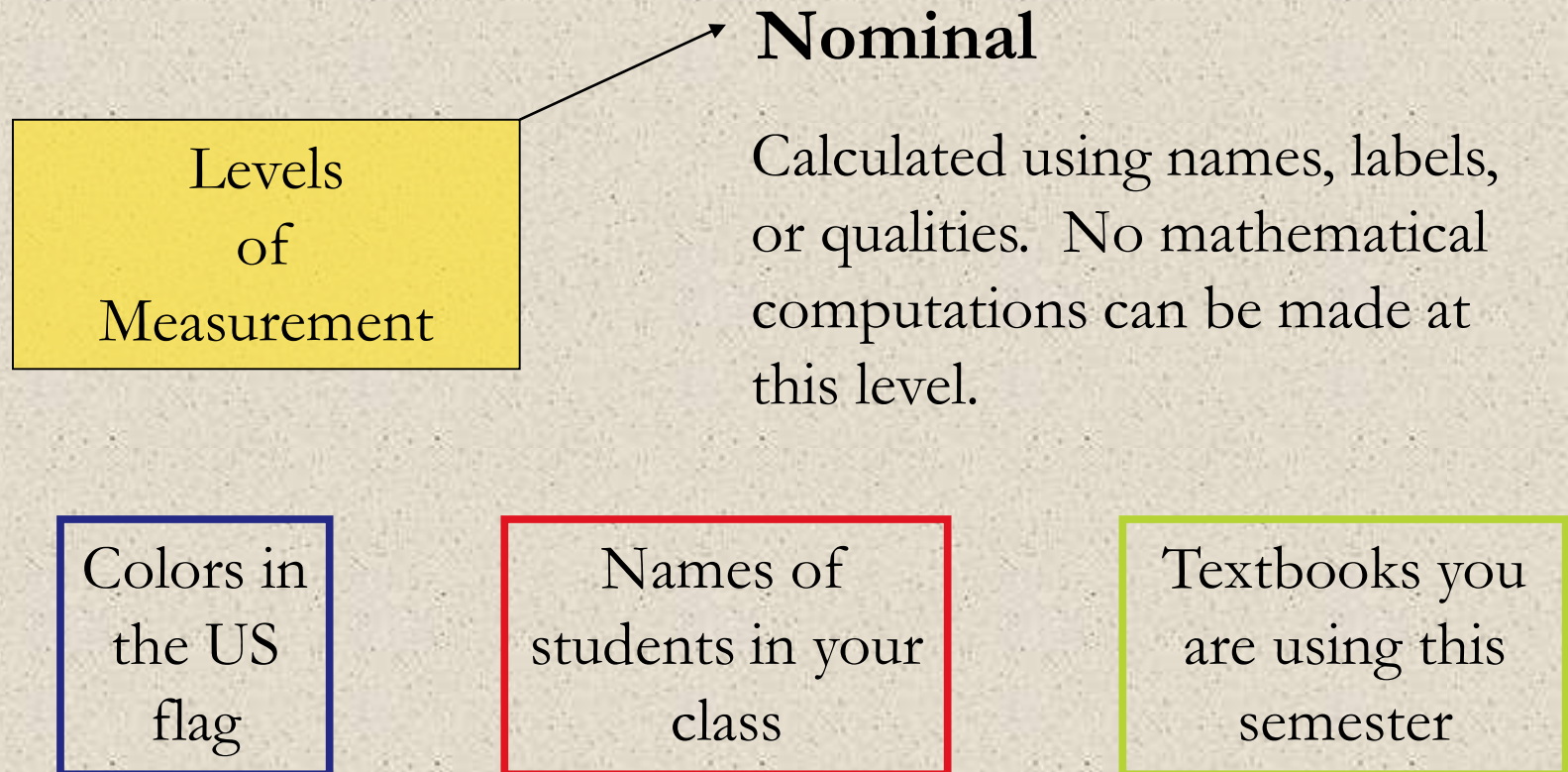
# Levels of Measurement

The level of measurement determines which statistical calculations are meaningful. The four levels of measurement are: **nominal**, **ordinal**, **interval**, and **ratio**.



# Nominal Level of Measurement

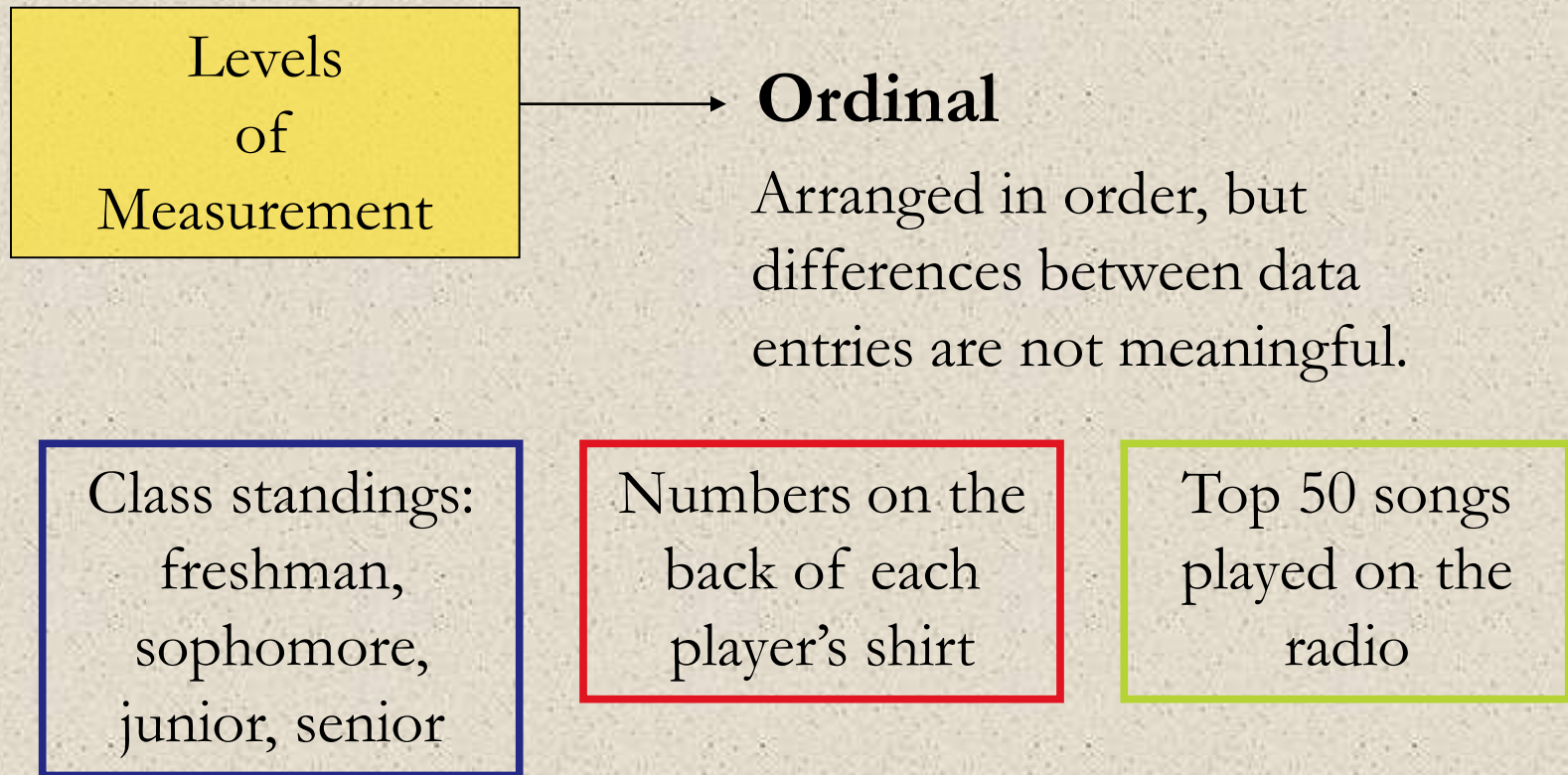
Data at the **nominal level of measurement** are qualitative only.





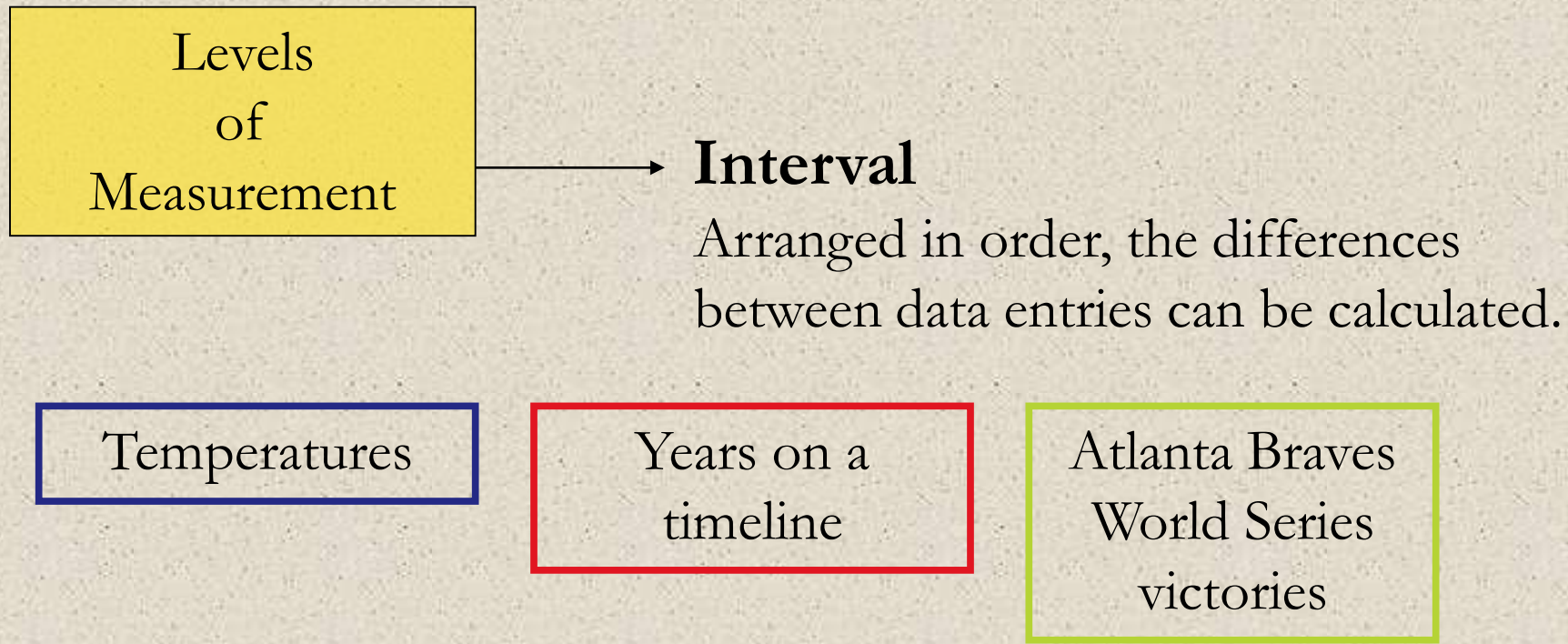
# Ordinal Level of Measurement

Data at the **ordinal level of measurement** are qualitative or quantitative.



# Interval Level of Measurement

Data at the **interval level of measurement** are quantitative. A zero entry simply represents a position on a scale; the entry is not an inherent zero.



# Ratio Level of Measurement

Data at the **ratio level of measurement** are similar to the interval level, but a zero entry is meaningful.

Levels  
of  
Measurement

A ratio of two data values can be formed so one data value can be expressed as a ratio.

**Ratio**

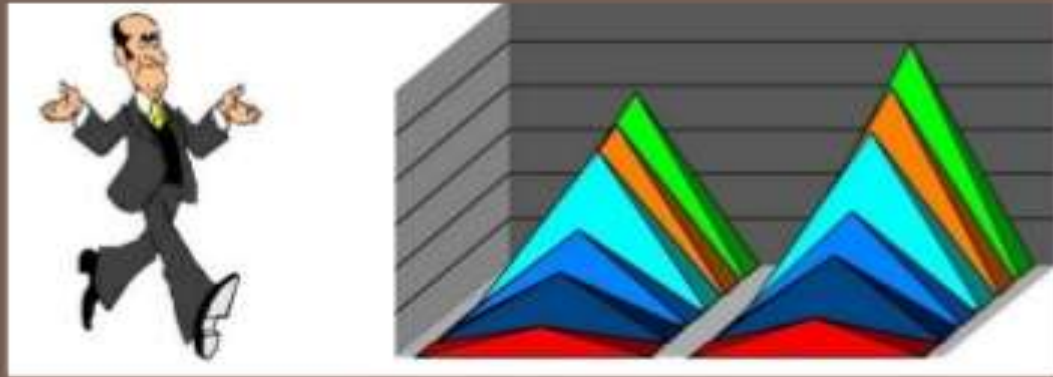
Ages

Grade point  
averages

Weights

# Summary of Levels of Measurement

Level of measurement	Put data in categories	Arrange data in order	Subtract data values	Determine if one data value is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes



# DATA COLLECTION PRIMARY & SECONDARY





# INTRODUCTION

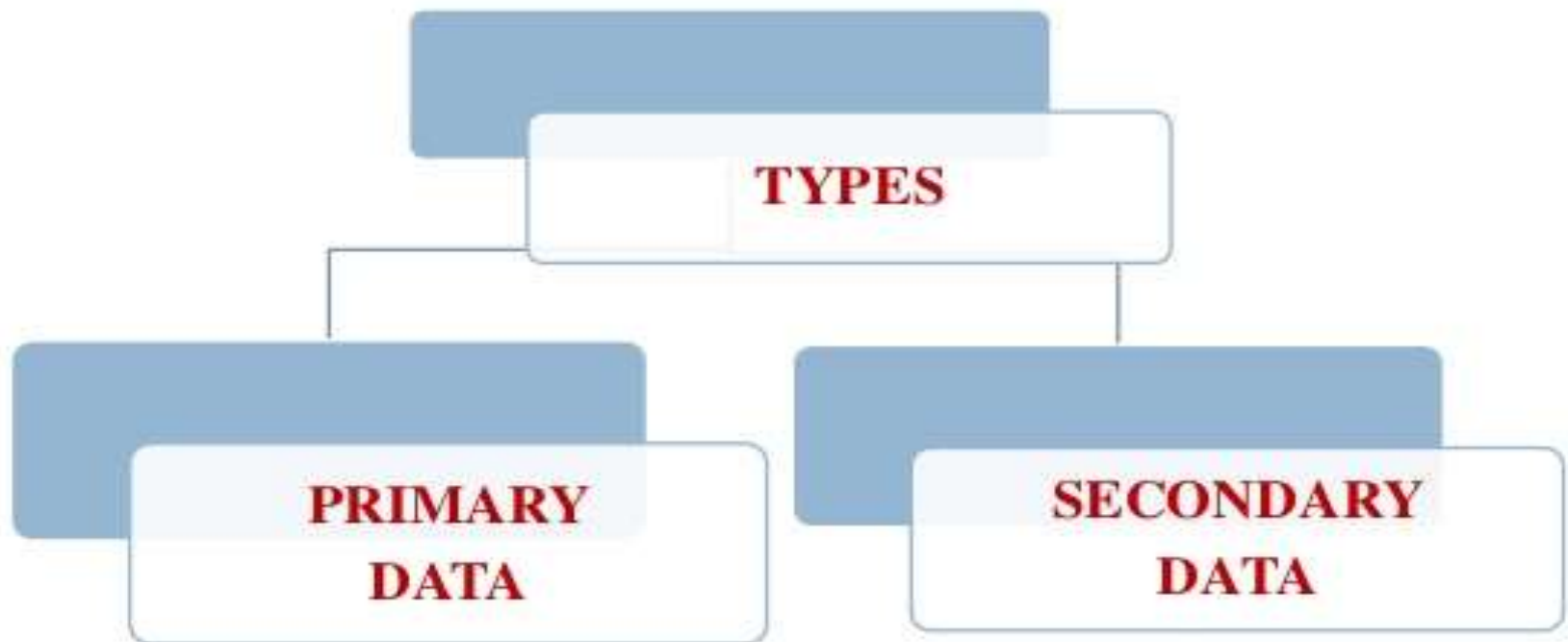
- ❑ Data collection is a term used to describe a process of preparing and collecting data
- ❑ Systematic gathering of data for a particular purpose from various sources, that has been systematically observed, recorded, organized.
- ❑ Data are the basic inputs to any decision making process in business

# PURPOSE OF DATA COLLECTION

- ❑ The purpose of data collection is-
  - ✓ to obtain information
  - ✓ to keep on record
  - ✓ to make decisions about important issues,
  - ✓ to pass information on to others



# CLASSIFICATION OF DATA



# PRIMARY DATA

- ❑ The data which are collected from the field under the control and supervision of an investigator
- ❑ Primary data means original data that has been collected specially for the purpose in mind
- ❑ This type of data are generally afresh and collected for the first time
- ❑ It is useful for current studies as well as for future studies
- ❑ For example: your own questionnaire.



# METHODS

- ❑ **OBSERVATION METHOD**  
Through personal observation
- ❑ **PERSONAL INTERVIEW**  
Through Questionnaire
- ❑ **TELEPHONE INTERVIEW**  
Through Call outcomes,  
Call timings
- ❑ **MAIL SURVEY**  
Through Mailed  
Questionnaire





# SECONDARY DATA

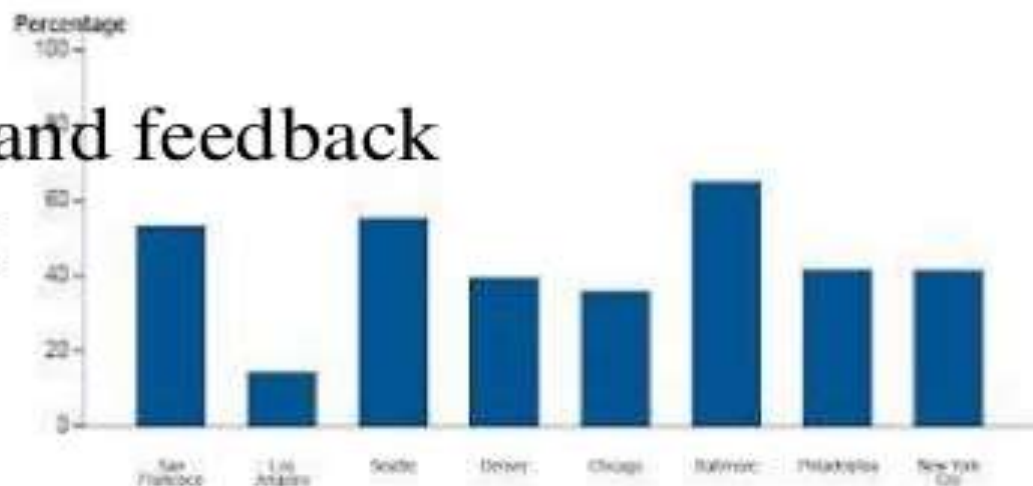
- ❑ Data gathered and recorded by someone else prior to and for a purpose other than the current project
- ❑ Secondary data is data that has been collected for another purpose.
- ❑ It involves less cost, time and effort
- ❑ Secondary data is data that is being reused. Usually in a different context.
- ❑ For example: data from a book.

# SOURCES

## □ INTERNAL SOURCES

Internal sources of secondary data are usually for marketing application-

- ✓ Sales Records
- ✓ Marketing Activity
- ✓ Cost Information
- ✓ Distributor reports and feedback
- ✓ Customer feedback



# SOURCES

## □ EXTERNAL SOURCES

External sources of secondary data are usually for Financial application-

- ✓ Journals
- ✓ Books
- ✓ Magazines
- ✓ Newspaper
- ✓ Libraries
- ✓ The Internet



# Variables

**Variable** is any characteristic of an individual.  
It can take on different values for different individuals



# Types of Variables

## Variables

### Quantitative

- Discrete
- Continuous

### Qualitative

- Ordinal
- Categorical

### Why is the type of variable important?

The methods used to display, summarize, and analyze data depend on whether the variables are categorical or quantitative.

# Quantitative Variables

- “A variable that quantifies an element of a population.”
- Numeric value that it makes sense to do arithmetic operations (+, -,  $\times$ , /)
- Finding an arithmetic “average” makes sense
- EX:
  1. Height
  2. Age
  3. Income
  4. Test Scores

# Two Types of Quantitative Variables

## **Discrete Variables**

It can assume a countable number of values.  
There is a gap between any two values.

## **Continuous Variables**

It can assume an uncountable number of values, i.e. can assume any value along a line interval, including every possible value between any two values.



# Discrete Random Variables

- A discrete random variable is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, 4,....
- Discrete random variables are usually (but not necessarily) counts.
- **Examples:**
  - number of children in a family
  - the Friday night attendance at a cinema
  - the number of patients a doctor sees in one day
  - the number of defective light bulbs in a box of ten
  - the number of “heads” flipped in 3 trials

# Continuous Random Variable

- A continuous random variable is one which takes an **infinite** number of possible values.
- Continuous random variables are usually measurements.
- Examples:
  - height
  - weight
  - the amount of sugar in an orange
  - the time required to run a mile.



**Variable** we already studied that it can be changed from object to object. Variables are denoted by capital alphabets X, Y, Z.....and so on.

**Constant** is a fix value which cannot be changed from object to object. Constant are denoted by small alphabets a, b, c, .....and so on.