A Course In Business Statistics

4th Edition

Chapter 1

The Where, Why, and How of Data Collection

Chapter Goals

After completing this chapter, you should be able to:

- Describe key data collection methods
- Know key definitions:
  - ♦ Population vs. Sample
  - ♦ Primary vs. Secondary data types
  - ♦ Qualitative vs. Qualitative data
  - ♦ Time Series vs. Cross-Sectional data
- Explain the difference between descriptive and inferential statistics
- Describe different sampling methods

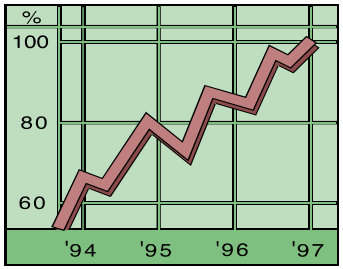
Tools of Business Statistics

- **Descriptive statistics**
  - Collecting, presenting, and describing data
- **Inferential statistics**
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Descriptive Statistics

- **Collect data**
  - e.g. Survey, Observation, Experiments

- **Present data**
  - e.g. Charts and graphs



- **Characterize data**
  - e.g. Sample mean = $\frac{\sum x_i}{n}$

Data Sources

Primary
Data Collection

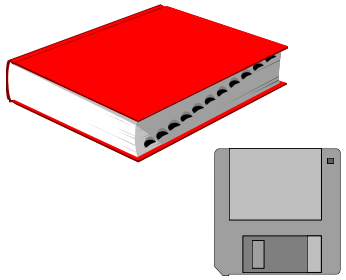
Observation

Survey

Experimentation

Secondary
Data Compilation

Print or Electronic



Survey Design Steps

- Define the issue
  - what are the purpose and objectives of the survey?
- Define the population of interest
- Formulate survey questions
  - make questions clear and unambiguous
  - use universally-accepted definitions
  - limit the number of questions

Survey Design Steps

(continued)

- Pre-test the survey
  - pilot test with a small group of participants
  - assess clarity and length
- Determine the sample size and sampling method
- Select Sample and administer the survey

Types of Questions

- **Closed-end Questions**
  - Select from a short list of defined choices

Example: Major: __business __liberal arts
__science __other

- **Open-end Questions**
  - Respondents are free to respond with any value, words, or statement

Example: What did you like best about this course?

- **Demographic Questions**
  - Questions about the respondents' personal characteristics

Example: Gender: __Female __ Male

Populations and Samples

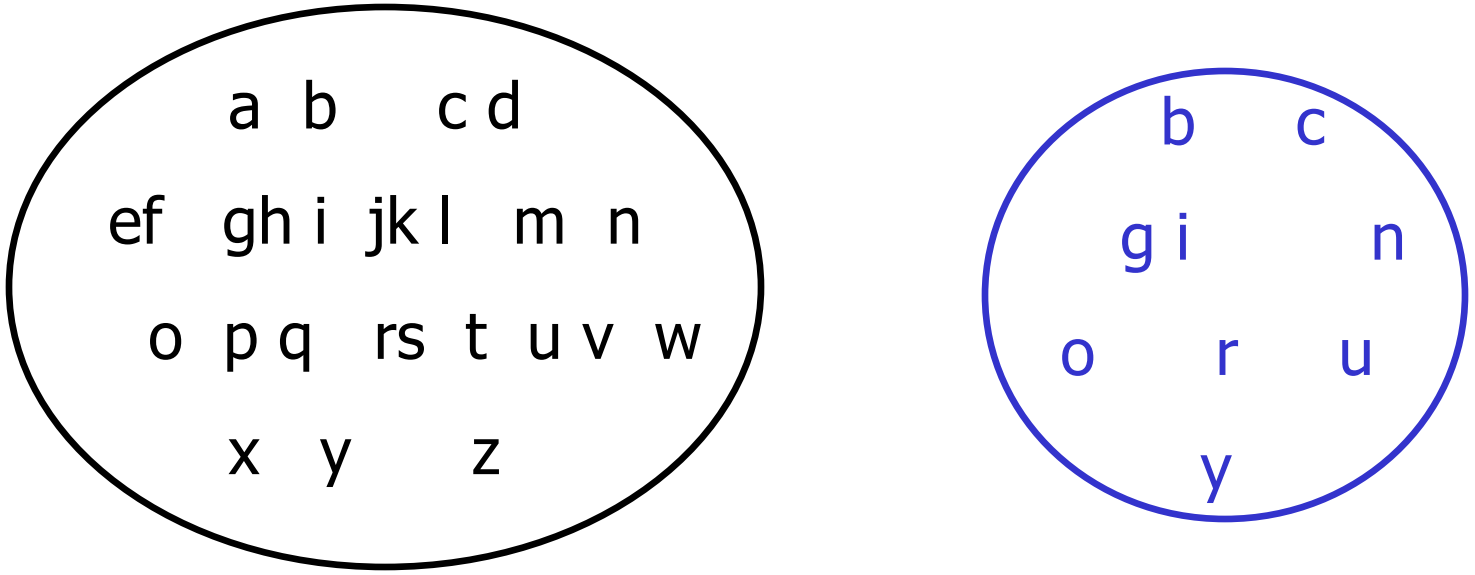
- A **Population** is the set of all items or individuals of interest

Examples:	All likely voters in the next election
	All parts produced today
	All sales receipts for November

- A **Sample** is a subset of the population

Examples:	1000 voters selected at random for interview
	A few parts selected for destructive testing
	Every 100th receipt selected for audit

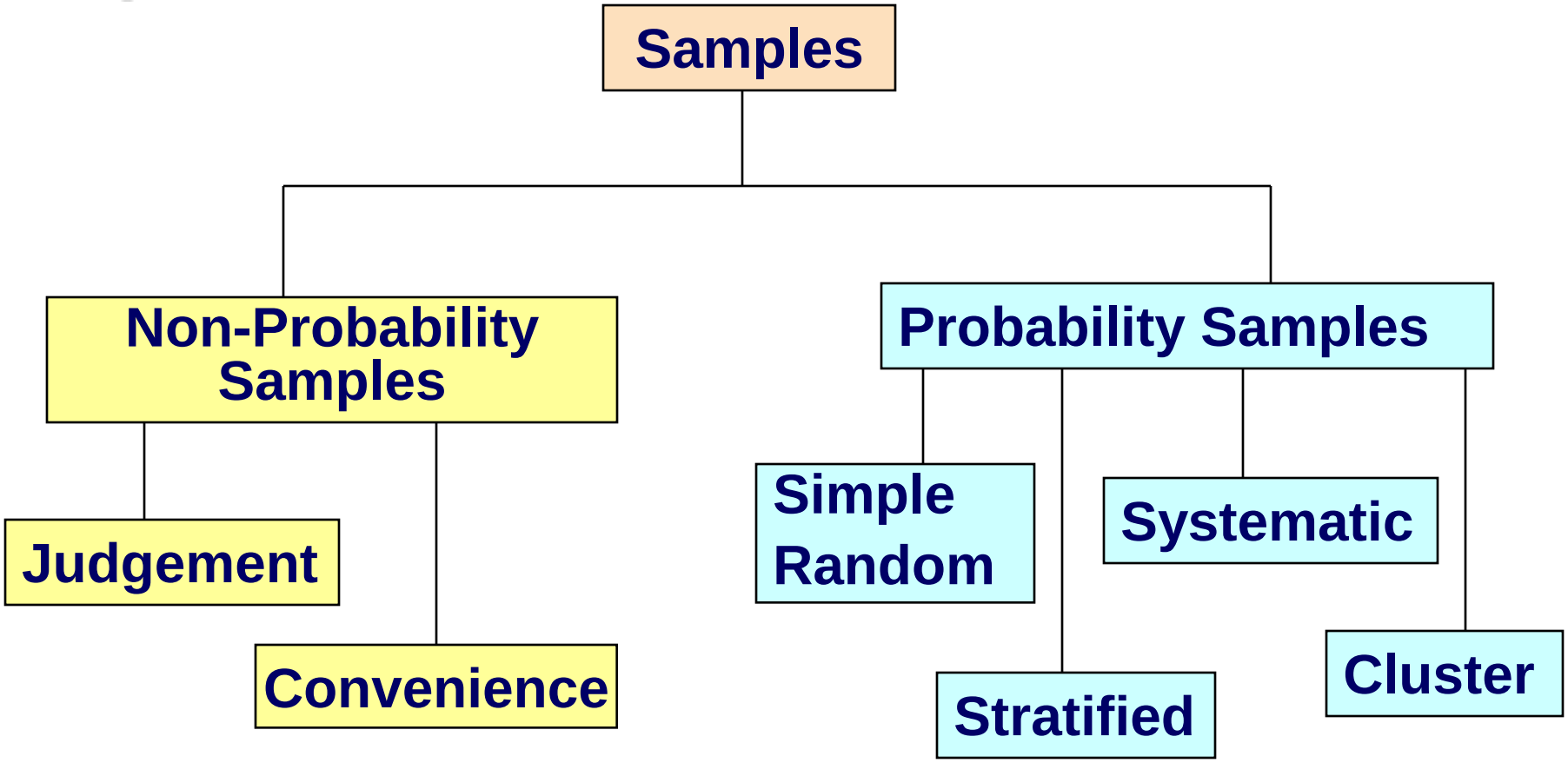
Population vs. Sample



Why Sample?

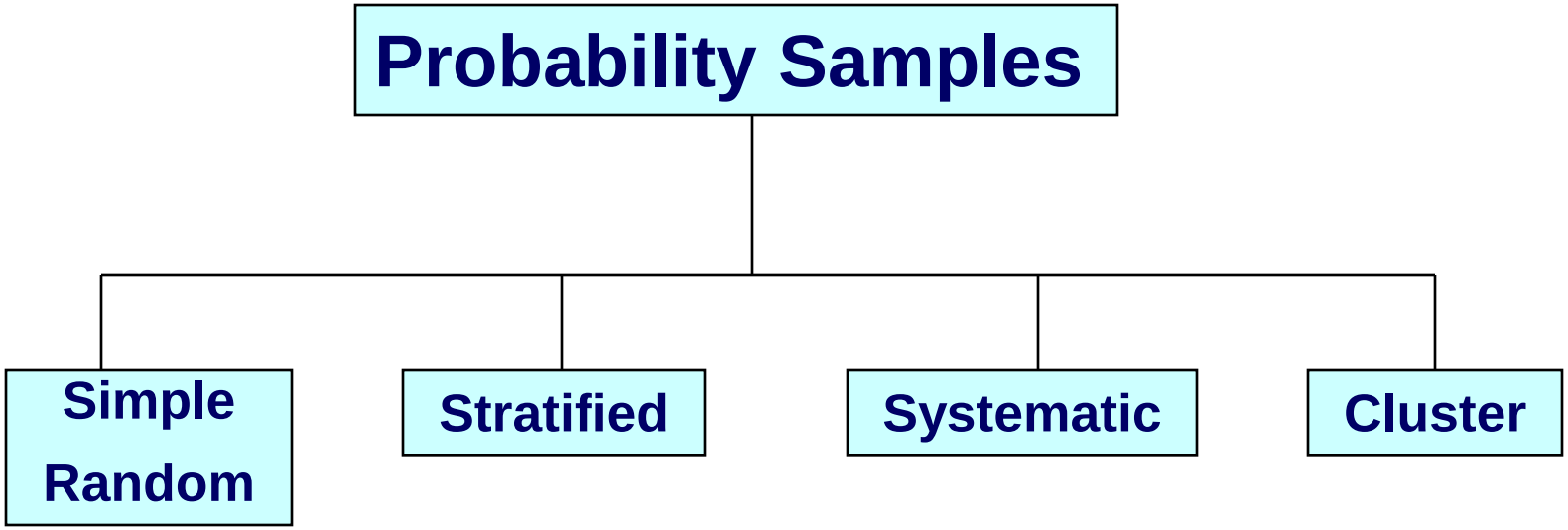
- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

Sampling Techniques



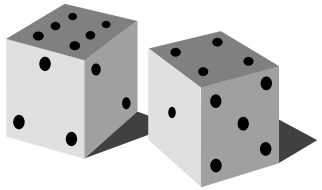
Statistical Sampling

- Items of the sample are chosen based on known or calculable probabilities



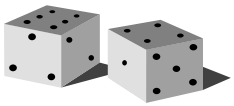
Simple Random Samples

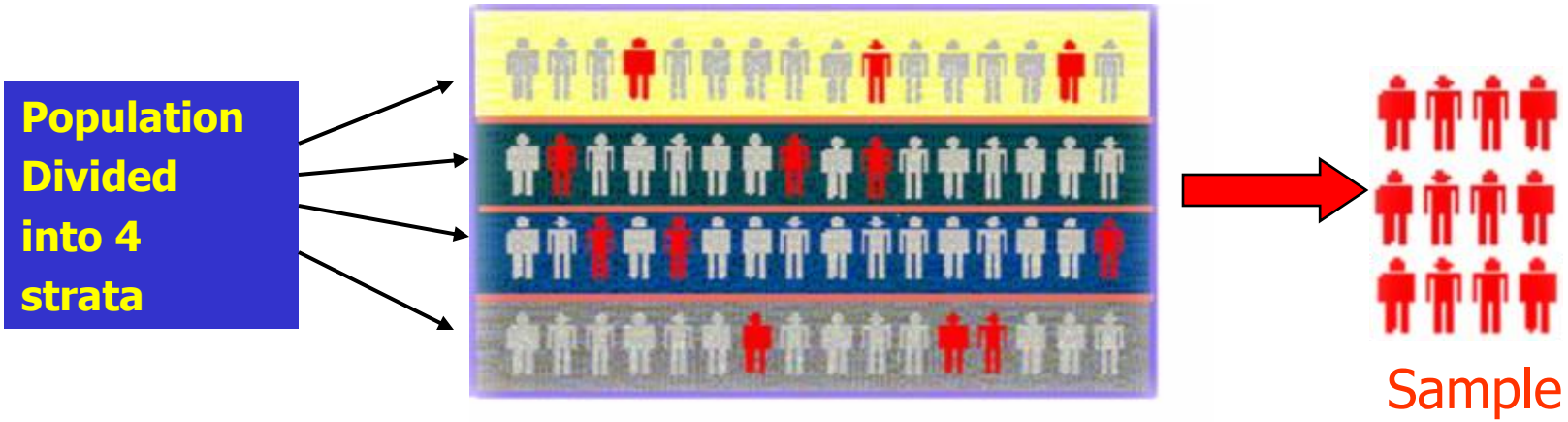
- Every individual or item from the population has an equal chance of being selected
- Selection may be with replacement or without replacement
- Samples can be obtained from a table of random numbers or computer random number generators



Stratified Samples

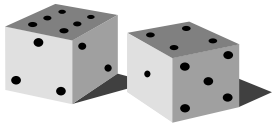
- Population divided into subgroups (called *strata*) according to some common characteristic
- Simple random sample selected from each subgroup
- Samples from subgroups are combined into one





Systematic Samples

- Decide on sample size: n
- Divide frame of N individuals into groups of k individuals: k=N/n
- Randomly select one individual from the 1st group



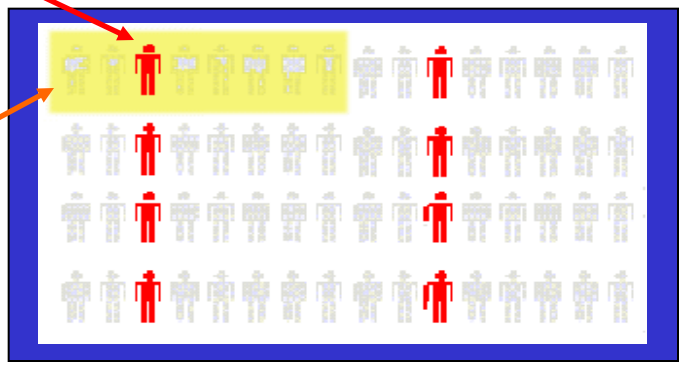
- Select every kth individual thereafter

N = 64

n = 8

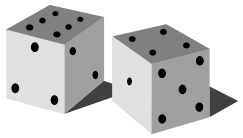
k = 8

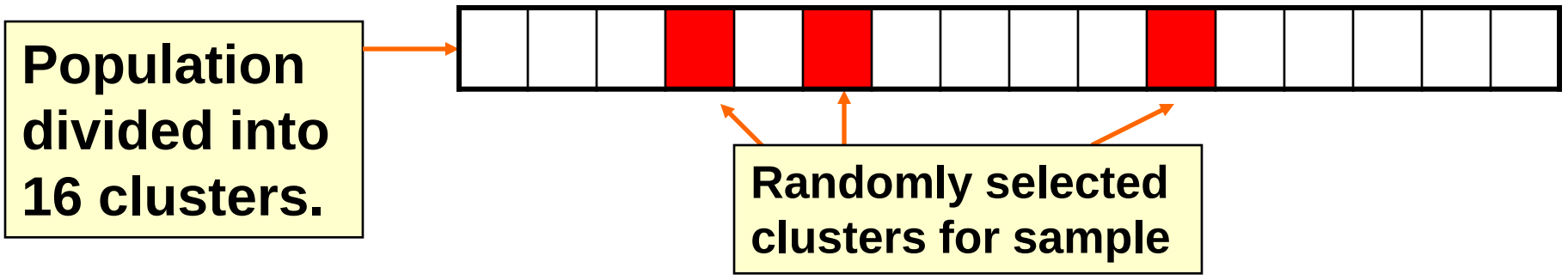
First Group



Cluster Samples

- Population is divided into several “clusters,” each representative of the population
- A simple random sample of clusters is selected
  - All items in the selected clusters can be used, or items can be chosen from a cluster using another probability sampling technique





Key Definitions

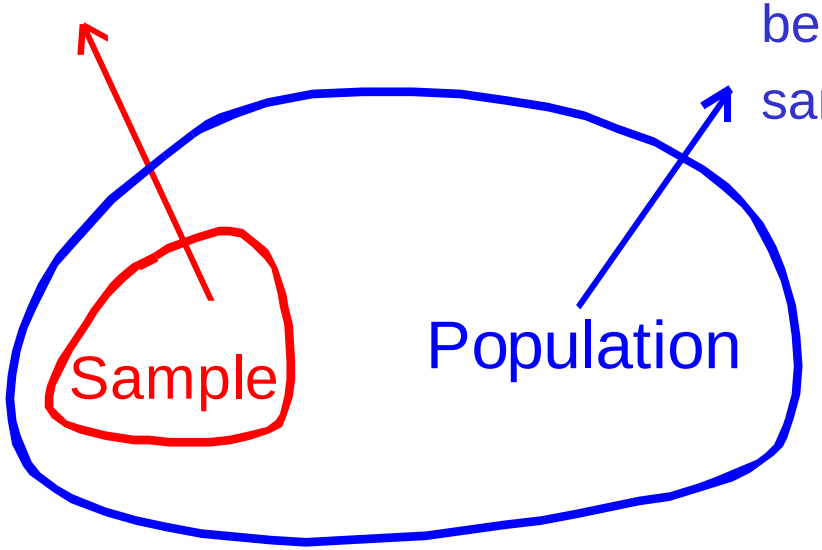
- A **population** is the entire collection of things under consideration
  - A **parameter** is a summary measure computed to describe a characteristic of the population
- A **sample** is a portion of the population selected for analysis
  - A **statistic** is a summary measure computed to describe a characteristic of the sample

Inferential Statistics

- Making statements about a population by examining sample results

Sample statistics → Population parameters

(known) **Inference** (unknown, but can be estimated from sample evidence)

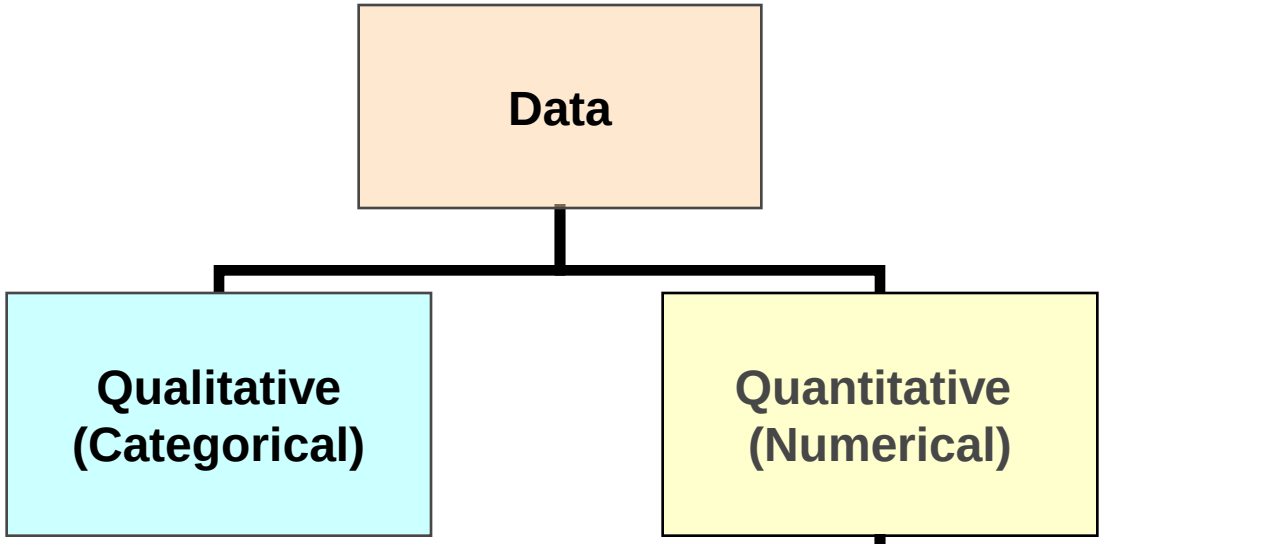


Inferential Statistics

Drawing conclusions and/or making decisions concerning a population based on sample results.

- **Estimation**
  - e.g.: Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g.: Use sample evidence to test the claim that the population mean weight is 120 pounds

Data Types



Examples:

- Marital Status
- Political Party
- Eye Color

(Defined categories)

Examples:

- Number of Children
- Defects per hour

(Counted items)

Examples:

- Weight
- Voltage

(Measured characteristics)

Data Types

- **Time Series Data**
  - Ordered data values observed over time
- **Cross Section Data**
  - Data values observed at a fixed point in time

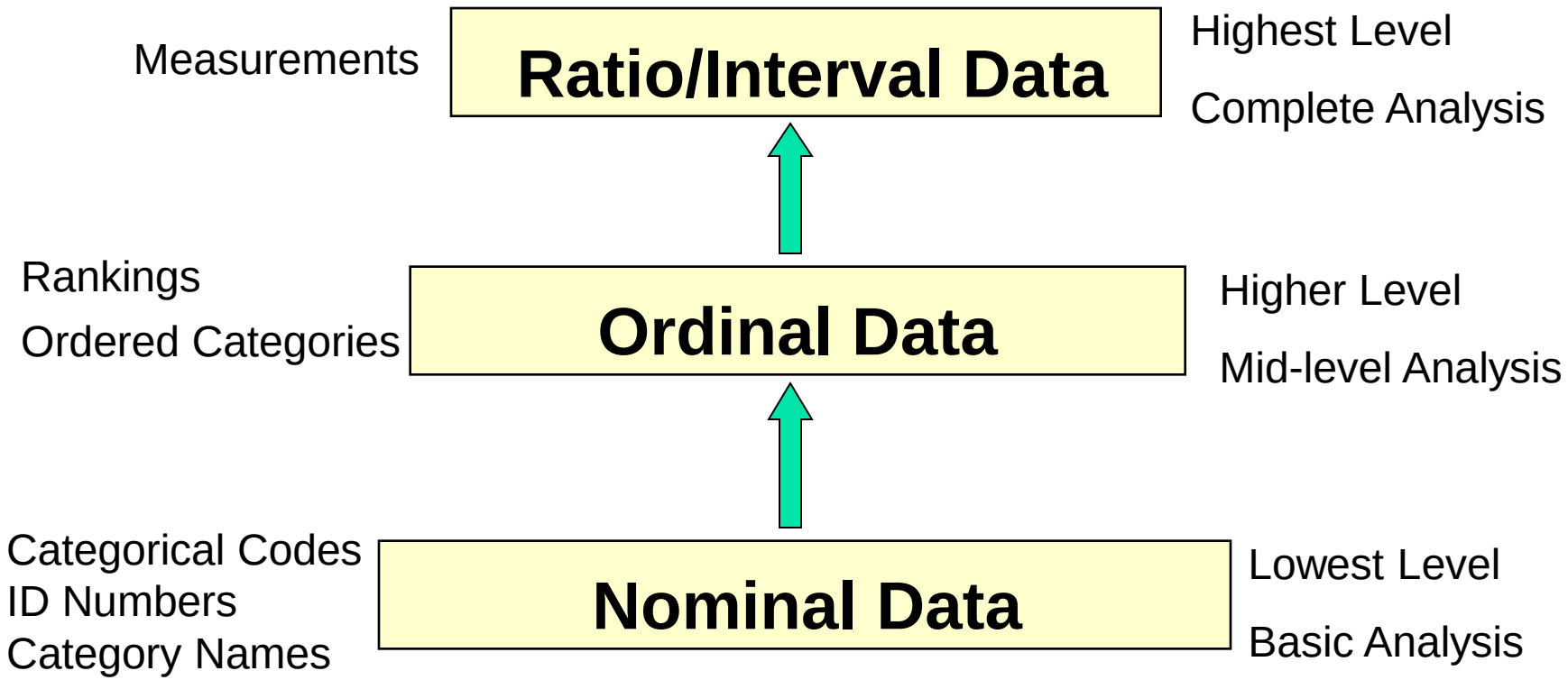
Data Types

	Sales (in $1000's)			
	2003	2004	2005	2006
Atlanta	435	460	475	490
Boston	320	345	375	395
Cleveland	405	390	410	395
Denver	260	270	285	280

Time Series Data

Cross Section Data

Data Measurement Levels



Chapter Summary

- Reviewed key data collection methods
- Introduced key definitions:
  - Population vs. Sample
  - Qualitative vs. Qualitative data
  - Primary vs. Secondary data types
  - Time Series vs. Cross-Sectional data
- Examined descriptive vs. inferential statistics
- Described different sampling techniques
- Reviewed data types and measurement levels