

# Chapter 2: Data Structures

## Classifying the Various Types of Data Sets

### Teaching Objectives

This material orients the student to the wide variety of data sets and the different questions and methods that are appropriate for each. Given a data set, the student should be able to tell (1) whether it is univariate, bivariate, or multivariate, (2) whether it is time-series or cross-sectional, (3) whether it is primary or secondary, and (4) whether it is an observational study or an experiment. Given a statistical variable, the student should be able to recognize it as quantitative (discrete or continuous) or qualitative (ordinal or nominal).

You may wish to stress the distinction between the elementary units (the objects being measured such as firms) and the variables (the type of measurement, such as sales dollars). It may help to observe that when the elementary units are successive days, then you will have a time-series. I tell my students “when you’re not sure what is represented in a data set, it will help you to identify the elementary units so that you can be sure just what the numbers represent: for example, it makes a difference whether these productivity measures are per person or per work group.”

Also mention the flexibility involved in these concepts. For example, a multivariate data set includes several univariate data sets as well as bivariate ones. Students should feel free to extract an interesting subset of a large data base and to view it as a data set in its own right.

The Internet is now a rich source of data, and can work to your advantage in motivating your students. For example, they may be able to locate data they are interested in, and then apply the chapter’s lesson (perhaps through the end-of-chapter projects) to a subject they genuinely care about.

### Question Answers

1. A data set consists of some basic measurement or measurements for each item of the analysis, the same piece or pieces of information being recorded for each item.
2. A variable is that piece of information recorded for each item being analyzed, its cost, for example.
3. The elementary units are the individual items or things which are measured. They may be people, trees or factories.

4.
  - a. How many pieces of information (variables) are there for each elementary unit? (Univariate, bivariate, or multivariate)
  - b. What kind of measurement (numbers or categories) is recorded in each case? (quantitative, qualitative, ordinal, nominal)
  - c. Is the time sequence of recording relevant or not? (Time series or cross-sectional data)
  - d. Did you control the data-gathering process or not? (Primary or secondary data)
  - e. Was a deliberate experiment involved to manipulate or control some characteristics, or were data merely observed? (An experiment or an observational data)
5.
  - a.
    - (1) What is a typical summary value?
    - (2) How diverse or different are these items?
    - (3) Are there any individuals or groups that require special attention?
  - b.
    - (1) Is there a simple relationship between the two variables?
    - (2) How strongly are they related?
    - (3) Can you predict one variable from the other? With what degree of reliability?
  - c.
    - (1) Is there a simple relationship between the variables?
    - (2) How strongly are they related?
    - (3) Can you predict one (a “special variable”) from the others? To what degree of reliability?
    - (4) Are there any individuals or groups that require special attention?
6. Analysis of bivariate data shows the relationship between the two univariate data sets.
7. With multivariate data, you can look at each variable individually, as well as examine the relationship among the variables and predict one variable from the others.
8. If the data come to you as meaningful numbers, then you have quantitative data. If the data set tells you which one of several non-numerical categories each item falls into, then they are called qualitative data (because they record some quality that the item possesses.)
9. A discrete variable can only take on values from a list of possible numbers, and no other values are conceivably possible. We will consider any numerical variable that is not discrete to be continuous. This word is used because typically the possible values form a “continuum” such as “any positive number” or “any number” or “any number between 0 and 100%.”

10. Qualitative data record some quality that the item possesses. The data set tells you which one of several non-numerical categories each item falls into.
11. A data set is ordinal if there is a meaningful ordering, so that you can speak of the first (perhaps the “best”), the second, the third, and so on. For nominal data you have only the categories, with no meaningful ordering; you have no meaningful numbers to compute with and no meaningful ordering to use for ranking.
12. If the data values are recorded in a meaningful sequence, such as daily stock market prices, then you have time series data. Otherwise, when the sequence in which the data are recorded is irrelevant, such as the first quarter 1987 earnings of 8 aerospace firms, you have cross-sectional data. With cross-sectional data a time sequence is not involved. For example, you might have a “cross-section” or snapshot of how things are at one particular time.
13. Analysis of time-series data is generally more complex than for cross-sectional data because the ordering of the observations must be carefully taken into account.
14. When you control the design of the data-collection plan (even if the work is done by others) you obtain **primary data**. When you use data previously collected by others for their own purposes, you are using **secondary data**. Primary data sets is often extensive and time-consuming to obtain, but can target exactly what you need. Secondary data sets are often inexpensive (or even free) but you might or might not find what you need.
15. In an **observational study** the data represent measurements as they occur naturally as part of the system being observed, while an **experiment** involves deliberate manipulation (such as randomization) to control some characteristic(s) of the system so that we can correctly assess what is causing a reaction of interest.

---

## Problem Solutions

---

1. Exercise for the student.
2. Exercise for the student.
3. Exercise for the student.
4. Exercise for the student.
5.
  - a. Exercise for the student.
  - b. Exercise for the student.
6. Exercise for the student.
7. Multivariate analysis could be used to predict one variable (your profit) based on several others (competitors’ performance, state of the economy, and time of year).

8.
  - a. Secondary (the government, not the company, collected the data).
  - b. Primary (data originated with the firm itself).
  - c. Secondary (data were purchased by, not collected by, your company).
9. This is a multivariate cross-sectional data set consisting of secondary data.
10. The elementary units of a telephone directory are people (and/or businesses, depending on the directory).
11.
  - a. The individual employee is the elementary unit for this data set.
  - b. This is a multivariate data set, with three or more columns.
  - c. Salary and years of experience are quantitative. Gender and education are qualitative.
  - d. Education is ordinal qualitative because the natural ordering HS, BA, MBA corresponds to more and more education.
  - e. These are cross-sectional data, without a natural sequence.
  - f. Observational study. These are simply measurements of the human-resources system at the time.
12.
  - a. Production facilities are the elementary units.
  - b. Multivariate.
  - c. Qualitative variables: part and quality. (Note that group ID identifies the elementary units and may or may not be viewed also as a qualitative variable).
  - d. Ordinal variable: quality.
  - e. Cross-sectional.
13.
  - a. Months are the elementary units.
  - b. Bivariate.
  - c. Both of these variables are quantitative.
  - d. Time-series.
14.
  - a. Customers are the elementary units.
  - b. Multivariate.
  - c. Quantitative variable: last year's total purchases. Qualitative variables: level of interest and geographical region.
  - d. Nominal: geographical region. Ordinal: level of interest.

- e. Cross-sectional.
15. Multivariate cross-sectional data. All variables are quantitative. Variables are (a) last year's spending for TV advertising, (b) last year's spending for radio advertising, and (c) last year's spending for newspaper advertising. Elementary units are competitors.
16. a. Time-series.  
b. Univariate.
17. a. Bivariate.  
b. Cross-sectional.
18. a. Univariate.  
b. Quantitative  
c. Neither.
19. a. Qualitative.  
b. Ordinal.
20. a. Qualitative.  
b. Nominal.
21. a. Models (of cell phones) are elementary units.  
b. Multivariate.  
c. Cross-sectional.  
d. Nominal.  
e. Quantitative.  
f. Ordinal.
22. Qualitative.
23. Ordinal.
24. Qualitative.
25. a. Vacuum cleaners are the elementary units.  
b. Multivariate.  
c. Quantitative: price and weight. Qualitative: quality and type.  
d. Quality is ordinal. Type is nominal.  
e. Cross-sectional.

26. a. Elementary units are companies.  
b. Bivariate.  
c. Both variables are quantitative.  
d. No variables are qualitative.  
e. Cross-sectional.
27. a. The elementary units are days.  
b. Multivariate.  
c. All variables are quantitative.  
d. No variables are qualitative.  
e. Time-series.
28. a. Observational study. We have measurements of the collection of customers who have joined the loyalty program, as part of this system.  
b. Experiment. There has been a deliberate manipulation (randomly choosing which large families receive free samples) which holds other factors equal so that we can discover causation, such as whether receiving free samples causes future orders to be larger.

---

### Database Exercise Solutions

---

1. a. Multivariate.  
b. The employees are the elementary units.  
c. Quantitative variables: salary, age, and experience. Qualitative variables: gender and training level.  
d. Training level is ordinal because it can be ranked in a meaningful way.  
e. No. This indicates that employee number is not a quantitative variable.  
f. Cross-sectional.
2. a. Appropriate for salary, age, and experience.  
b. All variables.  
c. Appropriate for salary, age, experience and training level.  
d. All variables.

\*\*\*\*\*

Data Sets from Chapter 2 of *Practical Business Statistics, Seventh Edition*

Copyright © 2016 by Andrew F. Siegel. Published by Elsevier, Inc. All Rights Reserved

\*\*\*\*\*

Profits, Profits as Percent of Stockholder Equity, Employees, and Revenues for Food Services Co

<u>Company</u>	millions		<u>Employees</u>	millions	
	<u>Profits</u>	<u>Profits Pct</u>		<u>Revenues</u>	
McDonald's	4,758	37%	420,000	27,441	
Starbucks	2,068	39%	191,000	16,448	
Yum Brands	1,051	67%	303,405	13,279	
Darden Restaurants	286	13%	206,489	8,758	
Bloomin' Brands	91	16%	100,000	4,443	
Chipotle Mexican Grill	445	22%	53,090	4,108	
Brinker International	154	244%	55,586	2,906	
Cracker Barrel Old Country Store	132	25%	72,000	2,684	
Panera Bread	179	24%	35,450	2,529	
Wendy's	121	7%	31,200	2,061	

Consumer Price Index - All Urban Consumers, 1-Month Percent Change

<u>Year Period</u>	<u>CPI</u>
2014 M01	0.4%
2014 M02	0.4%
2014 M03	0.6%
2014 M04	0.3%
2014 M05	0.3%
2014 M06	0.2%
2014 M07	0.0%
2014 M08	-0.2%
2014 M09	0.1%
2014 M10	-0.3%
2014 M11	-0.5%
2014 M12	-0.6%
2015 M01	-0.5%
2015 M02	0.4%
2015 M03	0.6%
2015 M04	0.2%
2015 M05	0.5%
2015 M06	0.4%

2015 M07

0.0%

## Regular Dividends from The Wall Street Journal

<u>Company</u>	<u>Dividend</u>
AbbVie Inc.	0.51
Amer Eagle Outfitters	0.125
Applied Materials	0.1
Cimarex	0.16
CST Brands	0.0625
General Electric	0.23
IDEX Corp	0.32
Intel	0.24
Liberty Property Tr	0.475
Shoe Carnival	0.065
Village Super Market Cl A	0.25

## Daily Stock Price Information for Home Depot

<u>Date</u>	<u>Open</u>	<u>High</u>	<u>Low</u>	<u>HD_Close</u>	<u>Volume</u>	<u>Adj Close</u>
8/28/2015	117.43	118.46	116.92	117.52	5002700	117.52
8/27/2015	117.5	118.09	115.33	117.66	7227000	117.66
8/26/2015	114.06	116.03	112.08	115.93	9298300	115.93
8/25/2015	117	117	110.89	110.97	10175700	110.97
8/24/2015	110.05	115.88	92.17	112.54	13628100	112.54
8/21/2015	119.54	120.36	116.15	116.16	10155000	116.16
8/20/2015	121.98	122.53	120.54	120.54	5549300	120.54
8/19/2015	122.67	123.79	122.13	122.76	6481700	122.76
8/18/2015	121.83	123.8	121.15	122.8	11015200	122.8
8/17/2015	119.1	120.75	118.73	119.7	7227700	119.7
8/14/2015	119.3	119.89	118.46	119.75	5133700	119.75
8/13/2015	118	120.14	117.79	119.02	6212600	119.02
8/12/2015	117.13	117.88	115.7	117.69	4575800	117.69
8/11/2015	117.02	118.14	116.76	117.69	3761900	117.69

Table 2.6.1. Employment/History Status of Five People.

<u>Gender</u>	<u>Salary</u>	<u>Education</u>	<u>Experience</u>
M	51,400	HS	13



F	56,200 BA	3
M	74,600 MBA	8
F	95,800 MBA	20
F	62,100 BA	11

Table 2.6.2. Selected Product Output of Production Facilities.

<u>Group Id</u>	<u>Part</u>	<u>Quality</u>	<u>Employ</u>
A-235-86	Brakes	good	53
W-186-74	Fuel line	better	37
X-937-85	Radio	fair	26
C-447-91	Chassis	excellent	85
F-258-89	Wire	good	16

Table 2.6.3. Sales and Income.

<u>Sales</u>	<u>Income</u>
350	30
270	23
140	(2)
280	14
410	53
390	47

Table 2.6.4. Selected Customers and Purchases.

<u>Product Interest</u>	<u>Purchases</u>	<u>Region</u>
Weak	88,906	West
Moderate	396,808	South
Very strong	438,442	South
Weak	2,486	Midwest
Weak	37,375	West
Very strong	2,314	Northeast
Moderate	1,244,096	Midwest
Weak	857,248	South
Strong	119,650	Northeast
Moderate	711,514	West
Weak	22,616	West

Table 2.6.5. Information about Cell Phones.

<u>Model</u>	<u>Operating System</u>	<u>Price</u>	<u>Screen Size</u>
Samsung Galaxy S5 4G LTE	Android KitKat	499	Small
Motorola Google Nexus 6	Android Lollipop	489	Large
LG G3 D855 4G LTE	Android KitKat	399	Medium
Apple iPhone 6	iOS 8	809	Small
Microsoft Lumia 640 XL RM-1096	Windows Phone	289	Large

Table 2.6.6. Comparison of Upright Vacuum Cleaners.

<u>Vacuum Price</u>	<u>Weight</u>	<u>Quality</u>	<u>Type</u>
170	17	good	hard-body
260	17	fair	soft-body, self-propelled
100	21	good	soft-body
90	14	good	hard-body
340	13	excellent	soft-body
120	24	good	soft-body, self-propelled
130	17	fair	soft-body, self-propelled

Table 2.6.7. Closing Price and Year-to-Date Percent Change for DJIA Firms.

	Closing Price 10/12/15	Percent Change From 01/02/15
<u>Company Name</u>	<u>Close</u>	<u>Change</u>
3M	150.06	-8.68%
American Express	77.31	-16.91%
Apple	111.60	1.11%
Boeing	140.68	8.23%
Caterpillar	70.50	-22.98%
Chevron	88.74	-20.89%
Cisco	27.96	0.52%
Coca-Cola	42.00	-0.52%
Disney	106.35	12.91%
E I du Pont de Nemours and Co	55.66	-20.76%
Exxon Mobil	79.30	-14.22%
General Electric	28.09	11.16%

Goldman Sachs	180.23	-7.02%
Home Depot	121.90	16.13%
IBM	151.14	-5.80%
Intel	32.21	-11.24%
Johnson & Johnson	95.99	-8.21%
JPMorgan Chase	61.72	-1.37%
McDonald's	103.24	10.18%
Merck	50.71	-10.71%
Microsoft	47.00	1.18%
Nike	126.43	31.49%
Pfizer	33.22	6.65%
Procter & Gamble	74.33	-18.40%
Travelers Companies Inc	103.72	-2.01%
United Technologies	95.43	-17.02%
UnitedHealth	122.51	21.19%
Verizon	44.30	-5.30%
Visa	74.99	14.40%
Wal-Mart	66.93	-22.07%

Table 2.6.8. Daily DJIA for September 2015.

<u>Date</u>	<u>DJIA</u>	<u>Net Change</u>	<u>Percent Change</u>
30-Sep-2015	16,284.70	235.57	1.47%
29-Sep-2015	16,049.13	47.24	0.30%
28-Sep-2015	16,001.89	-312.78	-1.92%
25-Sep-2015	16,314.67	113.35	0.70%
24-Sep-2015	16,201.32	-78.57	-0.48%
23-Sep-2015	16,279.89	-50.58	-0.31%
22-Sep-2015	16,330.47	-179.72	-1.09%
21-Sep-2015	16,510.19	125.61	0.77%
18-Sep-2015	16,384.58	-290.16	-1.74%
17-Sep-2015	16,674.74	-65.21	-0.39%
16-Sep-2015	16,739.95	140.10	0.84%
15-Sep-2015	16,599.85	228.89	1.40%
14-Sep-2015	16,370.96	-62.13	-0.38%
11-Sep-2015	16,433.09	102.69	0.63%
10-Sep-2015	16,330.40	76.83	0.47%
9-Sep-2015	16,253.57	-239.11	-1.45%
8-Sep-2015	16,492.68	390.30	2.42%
4-Sep-2015	16,102.38	-272.38	-1.66%
3-Sep-2015	16,374.76	23.38	0.14%

## Chap002

2-Sep-2015	16,351.38	293.03	1.82%
1-Sep-2015	16,058.35	-469.68	-2.84%
31-Aug-2015	16,528.03	--	--

mpanies.

HD Percent Change

-0.12%  
1.49%  
4.47%  
-1.40%  
-3.12%  
-3.63%  
-1.81%  
-0.03%  
2.59%  
-0.04%  
0.61%  
1.13%  
0.00%









