Statistical Analysis in Fin Mkts

MSF 502

Li Cai

CHAPTER 2

Tabular and Graphical Methods

Chapter 2 Learning Objectives (LOs)

LO 2.1: Summarize qualitative data by forming frequency distributions.

LO 2.2: Construct and interpret pie charts and bar charts.

LO 2.3: Summarize quantitative data by forming frequency distributions.

LO 2.4: Construct and interpret histograms, polygons, and ogives.

LO 2.5: Construct and interpret a stem-and-leaf diagram.

LO 2.6: Construct and interpret a scatterplot.

House Prices in Southern Californ

- A relocation specialist for a real estate firm in Mission Viejo, CA gathers recent house sales data for a client from Seattle, WA.
- The table below shows the sale price (in $1,000s) for 36 single-family houses.

$430	670	530	521	669	445
520	417	525	350	660	412
460	533	430	399	702	735
475	525	330	560	540	537
670	538	575	440	460	630
521	370	555	425	588	430

House Prices in Southern California

Use the sample information to:

1. Summarize the range of house prices.

2. Comment on where house prices tend to cluster.

3. Calculate percentages to compare house prices.

2.1 Summarizing Qualitative Data

LO 2.1 Summarize qualitative data by forming frequency distributions.

- **A frequency distribution for qualitative data groups data into categories and records how many observations fall into each category.**
- **Weather conditions in Seattle, WA during February 2010.**

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1 Rainy	2 Rainy	3 Rainy	4 Rainy	5 Rainy	6 Rainy
7 Rainy	8 Rainy	9 Cloudy	10 Rainy	11 Rainy	12 Rainy	13 Rainy
14 Rainy	15 Rainy	16 Rainy	17 Sunny	18 Sunny	19 Sunny	20 Sunny
21 Sunny	22 Sunny	23 Rainy	24 Rainy	25 Rainy	26 Rainy	27 Rainy
28 Sunny						

LO 2.1 2.1 Summarizing Qualitative Data

- **Categories: *Rainy*, *Sunny*, or *Cloudy*.**
- **For each category's frequency, count the days that fall in that category.**
- **Calculate relative frequency by dividing each category's frequency by the sample size.**

Weather	Frequency
Cloudy	1
Rainy	20
Sunny	7
Total	28

Weather	Frequency	Relative Frequency
Cloudy	1	1/28=0.036
Rainy	20	20/28=0.714
Sunny	7	7/28=0.250
Total	28	28/28=1.000

LO 2.1 2.1 Summarizing Qualitative Data

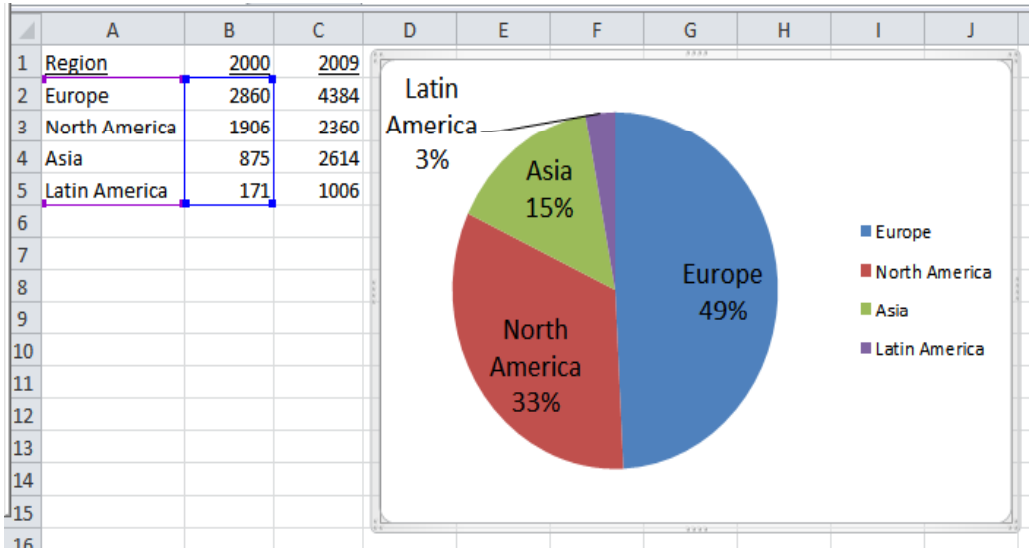
- **To express relative frequencies in terms of percentages, multiply each proportion by 100%.**

Weather	Frequency	Relative Frequency	Percentage
Cloudy	1	1/28=0.036	x 100= 3.6%
Rainy	20	20/28=0.714	x 100=71.4%
Sunny	7	7/28=0.250	x 100=25.0%
Total	28	28/28=1.000	x 100=100%

- **Note that the total of the proportions must add to 1.0 and the total of the percentages must add to 100%.**

LO 2.2 2.1 Summarizing Qualitative Data

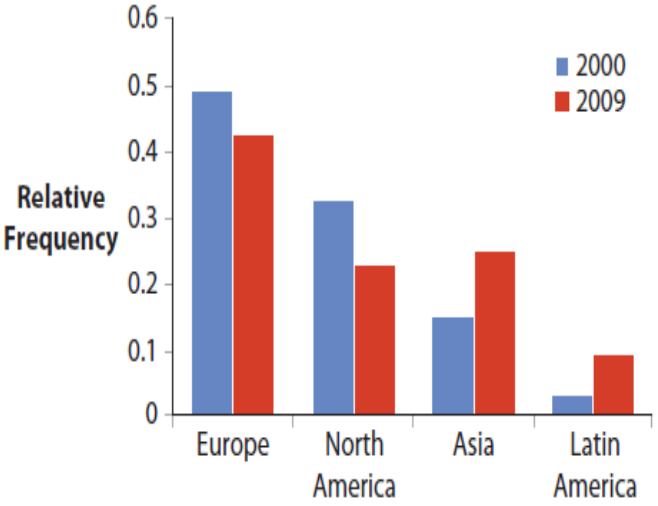
- **A pie chart is a segmented circle whose segments portray the relative frequencies of the categories of some qualitative variable.**
  - **In this example, the variable *Region* is proportionally divided into 4 parts.**



LO 2.2

2.1 Summarizing Qualitative Data

- **A bar chart depicts the frequency or the relative frequency for each category of the qualitative data as a bar rising vertically from the horizontal axis.**
  - **For example, Adidas' sales may be proportionally compared for each Region over these two periods.**



2.2 Summarizing Quantitative Data

LO 2.3 Summarize quantitative data by forming frequency distributions.

- **A frequency distribution for quantitative data groups data into intervals called classes, and records the number of observations that fall into each class.**
- **Guidelines when constructing frequency distribution:**
  - **Classes are *mutually exclusive*.**
  - **Classes are *exhaustive*.**

LO 2.3

2.2 Summarizing Quantitative Data

- **The number of classes usually ranges from 5 to 20.**

- **Approximating the class width:**

$$\frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}$$

LO 2.3

2.2 Summarizing Quantitative Data

- **The raw data from the Introductory Case has been converted into a frequency distribution in the following table.**

Class (in $1000s)	Frequency
300 up to 400	4
400 up to 500	11
500 up to 600	14
600 up to 700	5
700 up to 800	2
Total	36

LO 2.3

2.2 Summarizing Quantitative Data

Class (in $1000s)	Frequency
300 up to 400	4
400 up to 500	11
500 up to 600	14
600 up to 700	5
700 up to 800	2
Total	36

- **Question: What is the price range over this time period?**
  - **$300,000 up to $800,000**
- **Question: How many of the houses sold in the $500,000 up to $600,000 range?**
  - **14 houses**

LO 2.3

2.2 Summarizing Quantitative Data

- **A cumulative frequency distribution specifies how many observations fall below the upper limit of a particular class.**

Class (in $1000s)	Frequency	Cumulative Frequency
300 up to 400	4	4
400 up to 500	11	4 + 11 = 15
500 up to 600	14	4 + 11 + 14 = 29
600 up to 700	5	4 + 11 + 14 + 5 = 34
700 up to 800	2	4 + 11 + 14 + 5 + 1 = 36
Total	36	

- **Question: How many of the houses sold for less than $600,000?**
  - **29 houses**

LO 2.3

2.2 Summarizing Quantitative Data

- **A relative frequency distribution identifies the proportion or fraction of values that fall into each class.**

$$\text{Class relative frequency} = \frac{\text{Class frequency}}{\text{Total number of observations}}$$

- **A cumulative relative frequency distribution gives the proportion or fraction of values that fall below the upper limit of each class.**

LO 2.3

2.2 Summarizing Quantitative Data

- Here are the relative frequency and the cumulative relative frequency distributions for the house-price data.

Class (in $1000s)	Frequency	Relative Frequency	Cumulative Relative Frequency
300 up to 400	4	4/36 = 0.11	0.11
400 up to 500	11	11/36 = 0.31	0.11 + 0.31 = 0.42
500 up to 600	14	14/36 = 0.39	0.11 + 0.31 + 0.39 = 0.81
600 up to 700	5	5/36 = 0.14	0.11 + 0.31 + 0.39 + 0.14 = 0.95
700 up to 800	2	2/36 = 0.06	0.11 + 0.31 + 0.39 + 0.14 + 0.06 ≈ 1.0
Total	36	1.0	

LO 2.3

2.2 Summarizing Quantitative Data

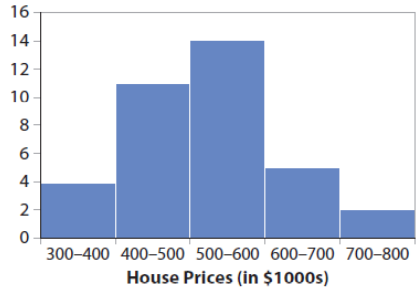
Use the data on the previous slide to answer the following two questions.

- **Question: What percent of the houses sold for at least $500,000 but not more than $600,000?**
  - **39%**
- **Question: What percent of the houses sold for less than $600,000?**
  - **81%**

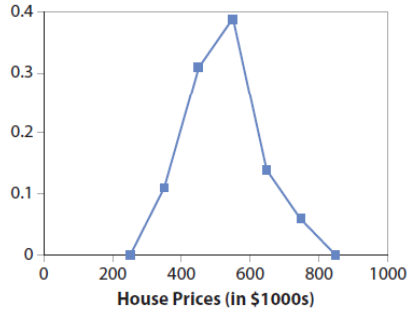
2.2 Summarizing Quantitative Data

LO 2.4 Construct and interpret histograms, polygons, and ogives.

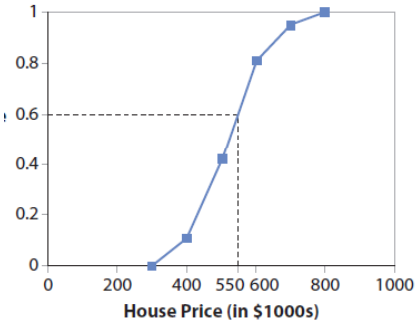
- **Histograms**



- **Polygons**



- **Ogives**



LO 2.4

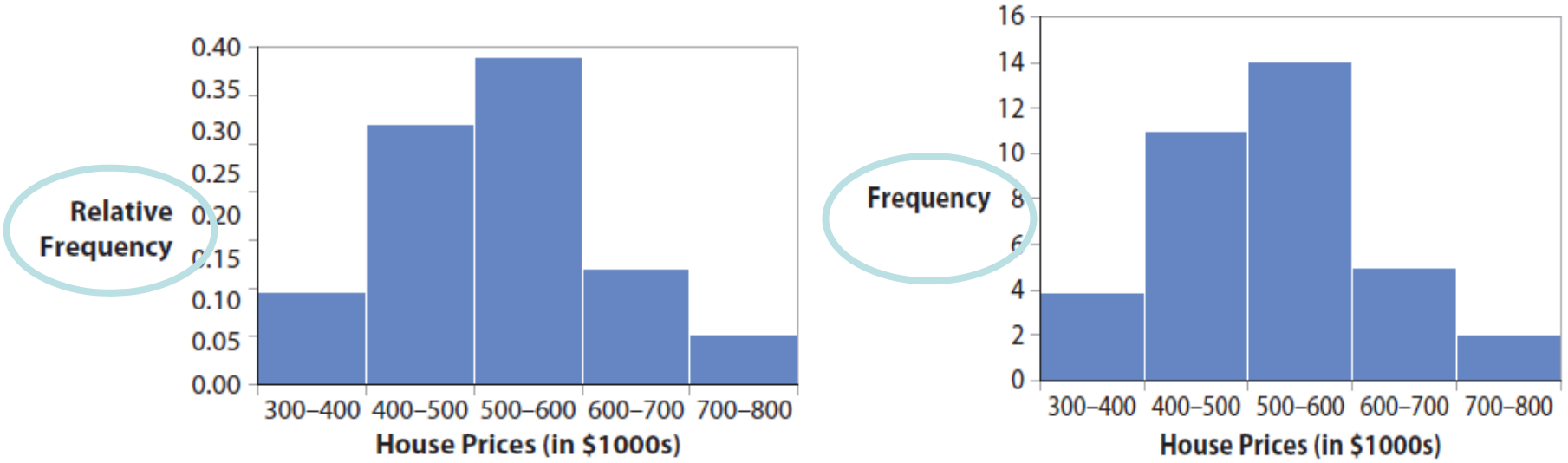
2.2 Summarizing Quantitative Data

- **A histogram is a visual representation of a frequency or a relative frequency distribution.**
  - **Bar height represents the respective class frequency (or relative frequency).**
  - **Bar width represents the class width.**

LO 2.4

2.2 Summarizing Quantitative Data

- **Here are the frequency and relative frequency histograms for the house-price data.**

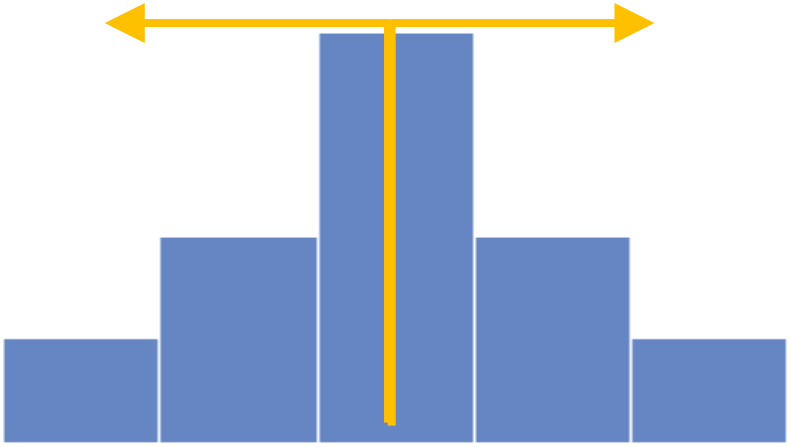


- **Note that the only difference is the *y*-axis scale.**

LO 2.4

2.2 Summarizing Quantitative Data

- **Shape of Distribution: typically symmetric or skewed**
  - **Symmetric—mirror image on both sides of its center.**

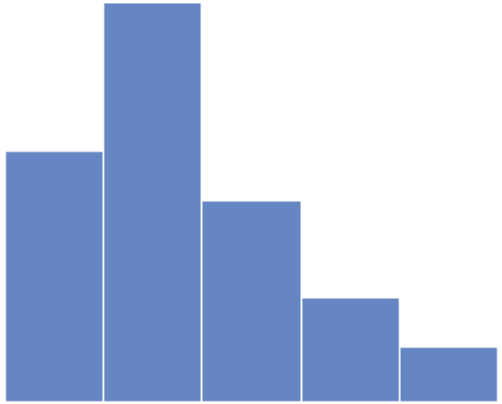


Symmetric Distribution

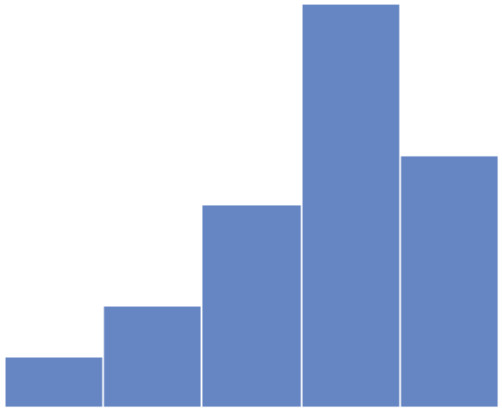
LO 2.4

2.2 Summarizing Quantitative Data

- **Skewed distribution**
  - **Positively skewed - data form a long, narrow tail to the right.**



  - **Negatively skewed - data form a long, narrow tail to the left.**



LO 2.4

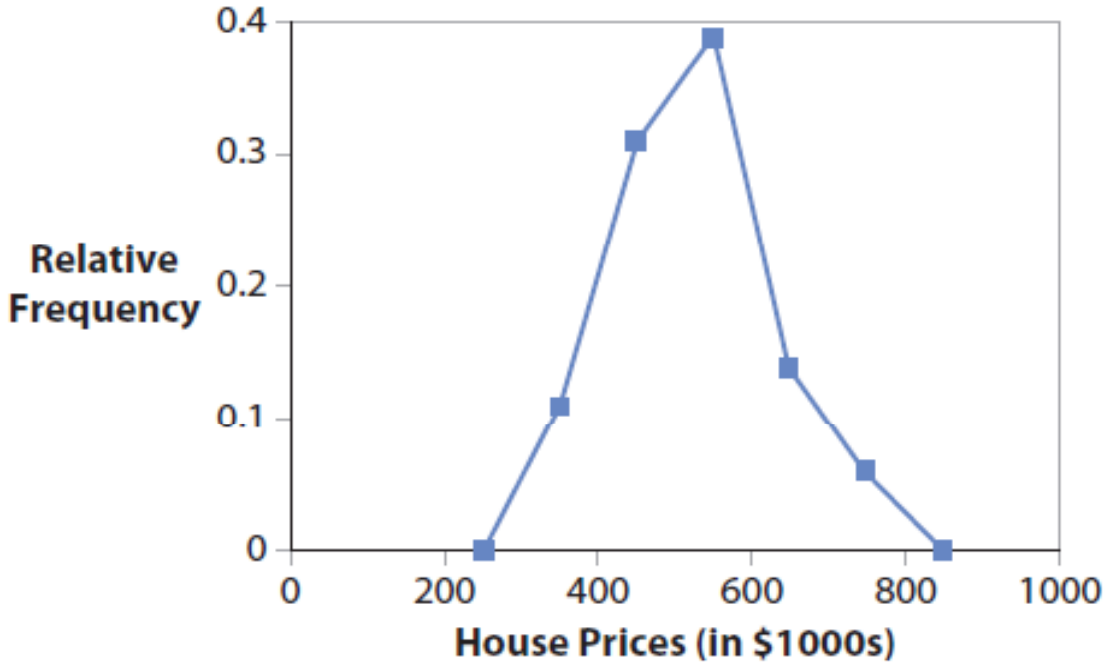
2.2 Summarizing Quantitative Data

- **A polygon is a visual representation of a frequency or a relative frequency distribution.**
  - **Plot the class midpoints on *x*-axis and associated frequency (or relative frequency) on *y*-axis.**
  - **Neighboring points are connected with a straight line.**

LO 2.4

2.2 Summarizing Quantitative Data

- **Here is a polygon for the house-price data.**



LO 2.4

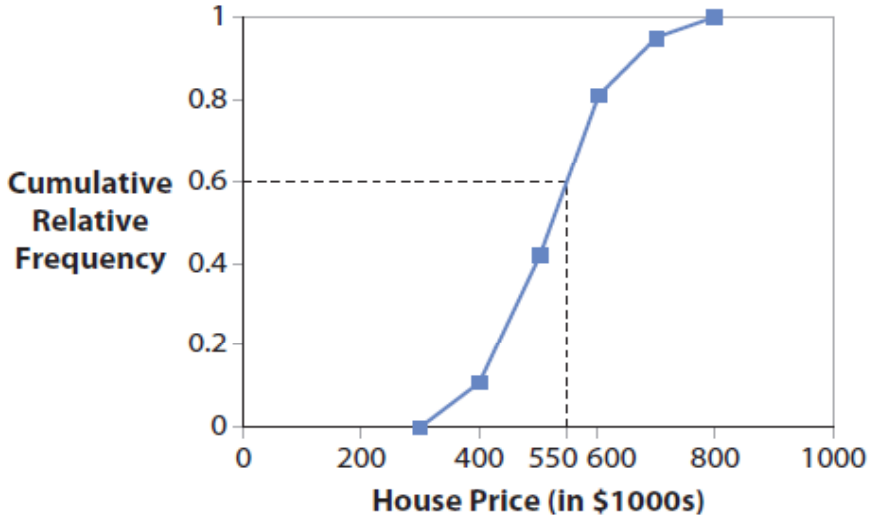
2.2 Summarizing Quantitative Data

- **An ogive is a visual representation of a cumulative frequency or a cumulative relative frequency distribution.**
  - **Plot the cumulative frequency (or cumulative relative frequency) of each class above the upper limit of the corresponding class.**
  - **The neighboring points are then connected.**

LO 2.4

2.2 Summarizing Quantitative Data

- **Here is an ogive for the house-price data.**



- **Use the ogive to approximate the percentage of houses that sold for less than $550,000.**

  Answer: 60%

2.3 Stem-and-Leaf Diagrams

LO 2.5 Construct and interpret a stem-and-leaf diagram.

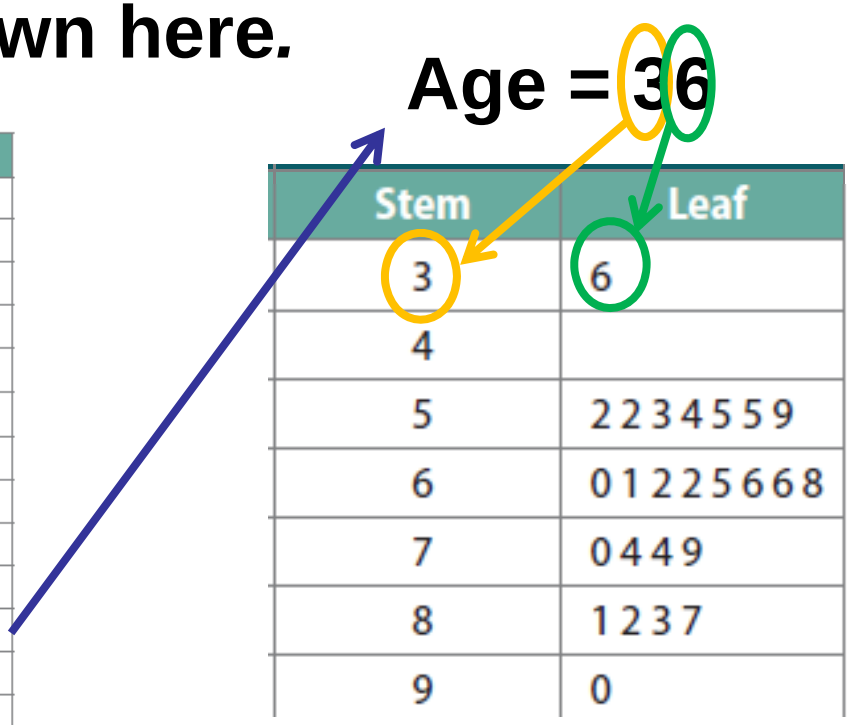
- **A stem-and-leaf diagram provides a visual display of quantitative data.**
- **It gives an overall picture of the data's center and variability.**
- **Each value of the data set is separated into two parts: the *stem* consists of the leftmost digits, while the *leaf* is the last digit.**

LO 2.5 2.3 Stem-and-Leaf Diagrams

- **The following data set shows the wealthiest people in the world and their associated ages.**
- **The leftmost digit is the *stem* while the last digit is the *leaf* as shown here.**

Name	Age	Name	Age
Carlos Slim Helu	70	Li Ka-shing	81
William Gates III	54	Jim Walton	62
Warren Buffet	79	Alice Walton	60
Mukesh Ambani	52	Liliane Bettencourt	87
Lakshmi Mittal	59	S. Robson Walton	66
Lawrence Ellison	65	Prince Alwaleed Alsaud	54
Bernard Arnault	61	David Thomson	52
Eike Batista	53	Michael Otto	66
Amancio Ortega	74	Lee Shau Kee	82
Karl Albrecht	90	Michael Bloomberg	68
Ingvar Kamprad	83	Sergey Brin	36
Christy Walton	55	Charles Koch	74
Stefan Persson	62		

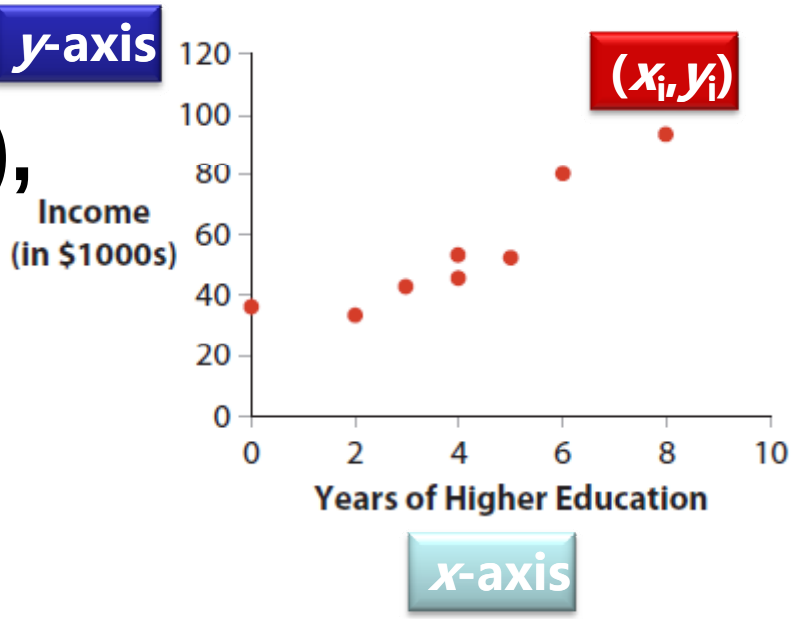
Source: www.forbes.com/lists/2010.



2.4 Scatterplots

LO 2.6 Construct and interpret a scatterplot.

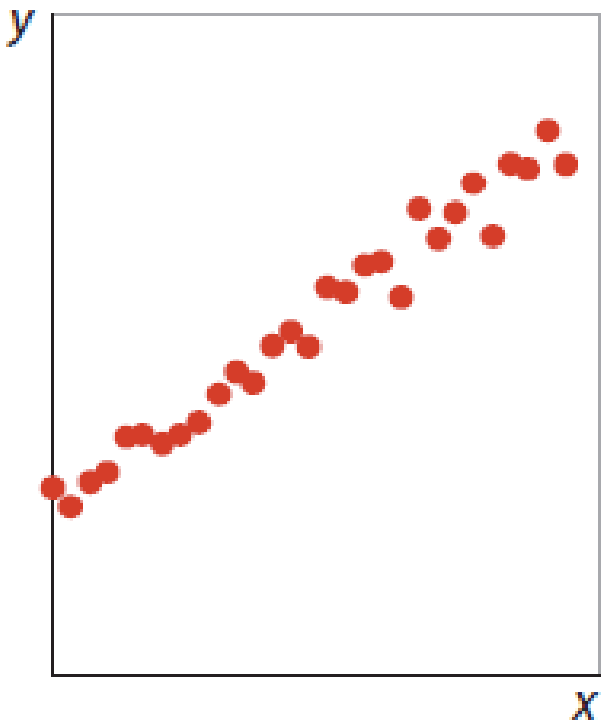
- **A scatterplot is used to determine if two variables are related.**
  - **Each point is a pairing: (x_1,y_1), (x_2,y_2), etc.**
  - **This scatterplot shows income against education.**



LO 2.6

2.4 Scatterplots

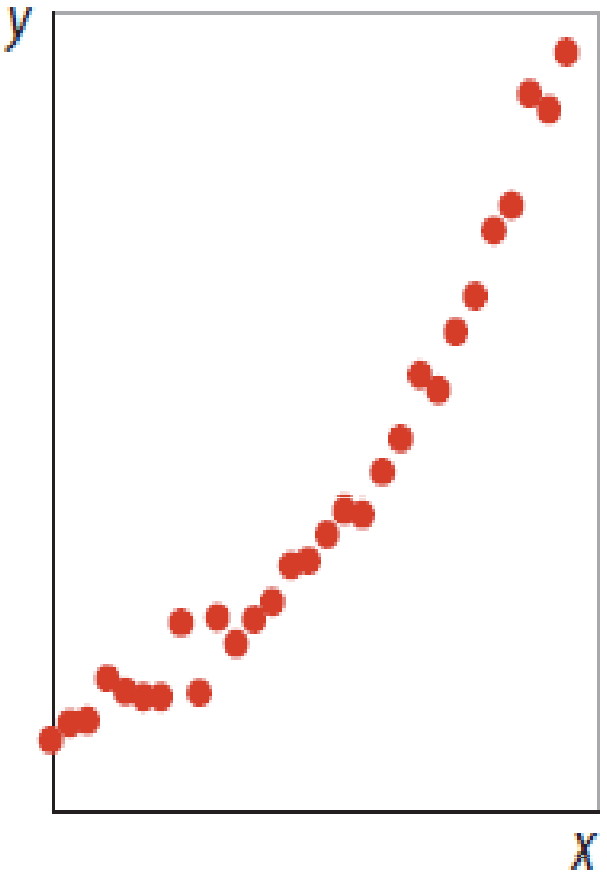
- **Linear relationship: upward or downward-sloping trend of the data.**
  - **Positive linear relationship (shown here): as *x* increases, so does *y*.**
  - **Negative linear relationship: as *x* increases, *y* decreases.**



LO 2.6

2.4 Scatterplots

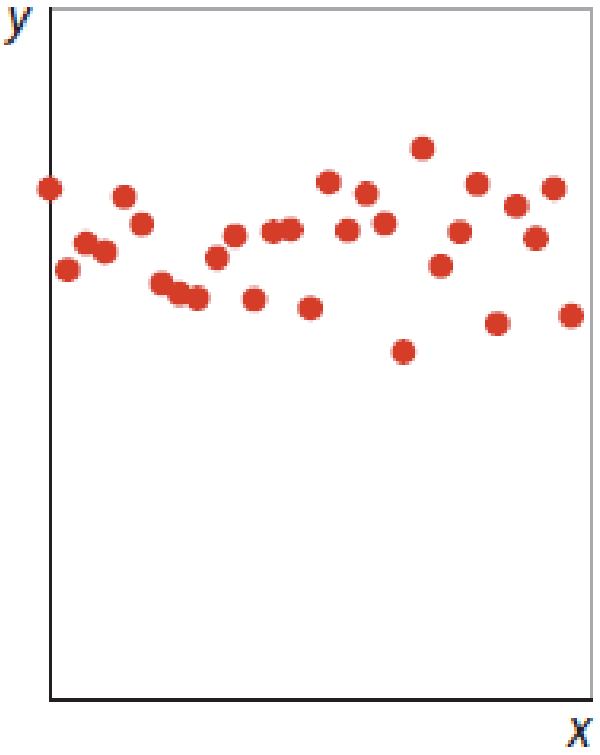
- **Curvilinear relationship**
  - **As *x* increases, *y* increases at an increasing (or decreasing) rate.**
  - **As *x* increases y decreases, at an increasing (or decreasing) rate.**



LO 2.6

2.4 Scatterplots

- **No relationship: data are randomly scattered with no discernible pattern.**
  - **In this scatterplot, there is no apparent relationship between *x* and *y*.**



LOs 2.1, 2.2, and 2.4

Some Excel Commands

- **Pie chart or Bar chart: select the relevant categorical names with respective data, then choose Insert > Pie > 2-D Pie or Insert > Bar > 2-D Bar.**

- **Histogram: select the relevant data, and choose Data > Data Analysis > Histogram.**

- **Scatterplot: select the *x*- and *y*-coordinates, choose Input > Scatter, and select the graph at the top left.**

End of Chapter