

Chapter 3: Graphical Descriptive Techniques II

Chapter 3: Graphical Descriptive Techniques II



The Histogram

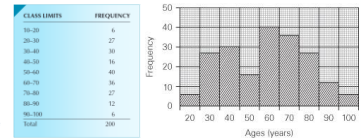
Ages of Duplicate Bridge Players

To learn more about the age of its players, the American Contract Bridge league (ACBL) collects a random sample of the age for 200 members.

DATA: [Xm03-01](#)

Solution:

The ages of the 200 members are grouped into contiguous and non-overlapping classes of 10 years each. Note that the first class includes age 10 but not age 20, the second class includes age 20 but not age 30, etc. so that every data point is included but never double-counted. Then, we can use the frequency of each class to build a histogram, which is a bar chart with the rectangles' height representing the frequency of each class, drawn adjacent to each other to convey continuity.



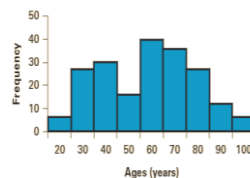
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage.

2

The Histogram with Data Analysis

Excel Instructions

1. Type or import the data from Xm03-01 into one column. In another column, type the upper limits of the class intervals. Excel calls them bins. Type "Ages" in the first row.
2. Click **Data**, **Data Analysis**, and **Histogram**.
3. Specify the **Input Range** (A1:A201) and the **Bin Range** (B1:B10). Click **Chart Output**. Click **Labels** if the first row contains names.
4. To remove the gaps, place the cursor over one of the rectangles and click the right button of the mouse. Click (with the left button) **Format Data Series**. Move the pointer to **Gap Width** and use the slider to change the number from 150 to 0.



Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

3

Determining the Number of Class Intervals

Two methods

NUMBER OF OBSERVATIONS	NUMBER OF CLASSES
Less than 50	5-7
50-200	7-9
200-500	9-10
500-1,000	10-11
1,000-5,000	11-13
5,000-50,000	13-17
More than 50,000	17-20

In Example 3.1, the table to the left suggests using at least 9 classes for a sample of size $n = 200$.

An alternative to the guidelines listed in the table to the left is to use **Sturges's formula**, which recommends that the number of class intervals be determined by the following:

$$\text{Number of class intervals} = 1 + 3.3 \log(n)$$

For $n = 200$, Sturges's formula becomes:

$$\begin{aligned} \text{Number of class intervals} &= 1 + 3.3 \log(200) \\ &= 1 + 3.3 \cdot 2.3 = 8.6 \end{aligned}$$

which we round to 9 classes.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

4

Chapter 3: Graphical Descriptive Techniques II

Determining the Class Interval Widths

We determine the approximate width of the classes by subtracting the smallest observation from the largest and dividing the difference by the number of classes:

$$\text{Class Width} = \frac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Number of Classes}}$$

In Example 3.1, we calculated:

$$\text{Class Width} = \frac{99 - 16}{9} = 9.22 \text{ years}$$

Then, we round the class width to 10 years, making sure that the first-class interval contains the smallest observation of age 16.

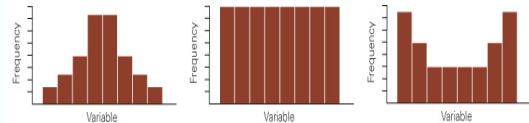
That's why in Example 3.1 we chose 9 classes of width 10 years each, beginning at age 10.

Keller, Gerald, Statistics for Management and Economics, 12th Edition, © 2023 Cengage. All Rights Reserved

5

Shapes of Histograms: Symmetry

Three Symmetric Histograms



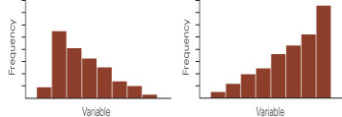
A histogram is said to be **symmetric** if, when we draw a vertical line down the center of the histogram, the two sides are identical in shape and size.

Keller, Gerald, Statistics for Management and Economics, 12th Edition, © 2023 Cengage. All Rights Reserved

6

Shapes of Histograms: Skewness

Positively and Negatively Skewed Histograms



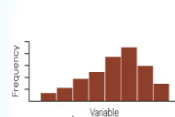
A **skewed** histogram has one of the two tails extending more than the other one. If the longer tail is to the right, the histogram is said to be **positively** skewed, and if the longer tail is to the left, it is said to be **negatively** skewed.

Keller, Gerald, Statistics for Management and Economics, 12th Edition, © 2023 Cengage. All Rights Reserved

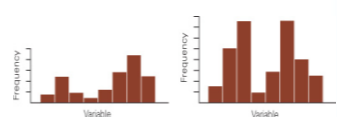
7

Shapes of Histograms: Number of Modal Classes

A unimodal histogram



Bimodal histograms



A **modal class** is the class with the largest number of observations.

A **unimodal** histogram is one with a single peak.

A **bimodal** histogram is one with two peaks, not necessarily equal in height. Bimodal histograms often indicate that two different distributions are present

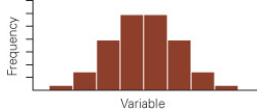
Keller, Gerald, Statistics for Management and Economics, 12th Edition, © 2023 Cengage. All Rights Reserved

8

Chapter 3: Graphical Descriptive Techniques II

Shapes of Histograms: Bell Shape

Bell-Shaped Histogram



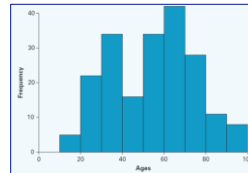
A **bell-shaped** histogram is a special type of symmetric unimodal histogram. In Chapter 8, we will explain why this type of histogram is important.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

9

Histograms - Stata

Example 3.1 – Stata Output and Instructions



Instructions:

1. Import or type the data into one column. (Click File/Import/Excel spreadsheet [Xm03-01](#).) Check **Import first row as variable names**.
2. Click **Graphics**, and **Histogram**.
3. Select **Ages** in the **Variable:** box. Check **Width of bins** and type 10. Check **Lower limit of first bin** and type 10. Select **Frequency**.

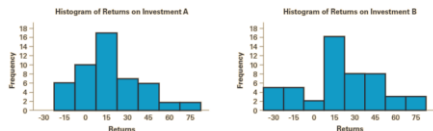
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

10

Comparing Returns on Two Investments

The returns for the two types of investments are listed here. Draw histograms for each set of returns and report on your findings. Which investment would you choose and why?

DATA: [Xm03-02](#)



Comparing the two histograms, we can extract the following information:

1. The center of the histogram of the returns of investment A is slightly lower than that for investment B.
2. The spread of returns for investment A is considerably less than that for investment B.
3. Both histograms are slightly positively skewed.

These findings suggest that investment A is superior.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

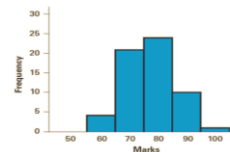
11

Business Statistics Marks

Concerned about a course's alleged difficulty, a student asks the professor about last year's marks.

The professor obliges and provides a list of the final marks, which is composed of term work plus the final exam.

Draw a histogram and describe the result, based on the marks included in file [Xm03-03+](#).



The histogram is unimodal and approximately symmetric.

There are no marks below 50, with the great majority of marks between 60 and 90.

The modal class is 70 to 80, and the center of the distribution is approximately 75.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

12

Chapter 3: Graphical Descriptive Techniques II

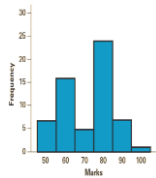
Mathematical Statistics Marks

Suppose the student in Example 3.3 obtained a list of last year's marks in a mathematical statistics course (DATA: [Xm03-03.xls](#)).

This course emphasizes derivations and proofs of theorems.

Use the accompanying data to draw a histogram and compare it to the one produced in Example 3.3.

What does this histogram tell you?



The histogram is bimodal, suggesting that there are two groups of students.

The larger modal class is composed of the marks in the 80s, and the smaller modal class includes marks in the 50s. There appear to be few marks in the 70s.

The histograms in this example and in Example 3.3 suggest that the courses are quite different from one another and have a completely different distribution of marks.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

13

Describing Time Series Data

Observations measured at the same point in time are called **cross-sectional data**.

Observations measured at successive points in time are called **time-series data**.

Time-series data are graphed on a **line chart**, which plots the value of the variable on the vertical axis against the time periods on the horizontal axis.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

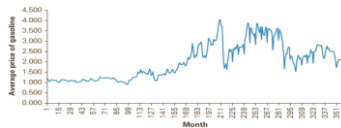
14

Price of Gasoline

Data file [Xm03-05](#) contains the monthly average retail price of gasoline (regular unleaded in dollars per gallon) since 1991.

Draw a line chart to visualize these figures and describe the results. We can use Excel or Stata

•



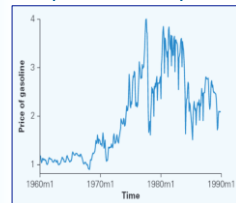
The price of gasoline rose from about \$1.00 to over \$4.00 in 2008 (around month 200) then dropped precipitously in March 2009 (at the end of the housing meltdown) before rising again and then falling to about \$2.00.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

15

Line Charts - Stata

Example 3.5 – Stata Output and Instructions



Instructions:

1. Import or type the data into one column. (Click File/Import/Excel spreadsheet (*.xls, *.xlsx)/[Xm03-05](#).) Check **Import first row as variable names**. One column must be a time variable. In this data set the variable is Time.
2. Click **Statistics, Time series, Setup and utilities**, and **Declare data set to be time-series data**. Select **Time** as Time variable. Check **Monthly**. Click **OK**.
3. Click **Graphics, Time-series graphs, and Line plots**.
4. Click **Create** and in the **Y variable** box select **Price of Gasoline**. Click **Time-series plot**. Click **Accept**. Click **OK**.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

16

Chapter 3: Graphical Descriptive Techniques II

Price of Gasoline in Constant Dollars

Data file [Xm03-06](#) contains the inflation-adjusted monthly average retail price of gasoline (regular unleaded in dollars per gallon) since 1991. Draw a line chart to visualize these figures and describe how gas price has changed after adjusting for inflation.



Using constant 1982–1984 dollars, we can see that the average price of a gallon of gasoline hit its peak in the middle of 2008 (around month 200). From there, it dropped rapidly and then rose and fell. By 2020 the adjusted price was lower than in January 1991.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

17

17

Describing the Relationship between Two Interval Variables

Moving from nominal data to interval data, we are frequently interested in how two interval variables are related.

To explore this relationship, we employ a **scatter diagram**, which plots two variables against one another.

The **independent** variable is labeled X and is usually placed on the horizontal axis, while the other, **dependent** variable, Y, is mapped to the vertical axis.

Observations measured at the same point in time are called **cross-sectional data**.

Observations measured at successive points in time are called **time-series data**.

Time-series data are graphed on a **line chart**, which plots the value of the variable on the vertical axis against the time periods on the horizontal axis.

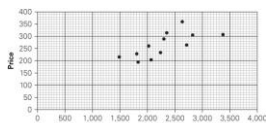
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

18

18

Relationship between Price and Size of House

A real estate agent recorded the price and size for a sample of 12 recently sold houses to analyze the relationship between the selling price of a home (\$1,000s) and its size (sqft). Data file: [Xm03-07](#). Describe the relationship between size and price with a scatter diagram.



The scatter diagram reveals that the greater the size of the house, the greater the price. However, there are other variables that determine price. Further analysis may reveal what these other variables are.

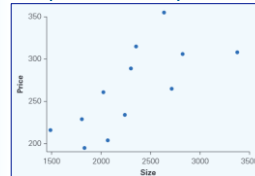
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

19

19

Scatter Diagram - Stata

Example 3.7 – Stata Output and Instructions



Instructions:

1. Import or type the data into one column. (Click File/Import/Excel spreadsheet (*.xls, *.xlsx)/[Xm03-07](#).) Check **Import first row as variable names**.
2. Click **Graphics** and **Twoway graph (scatter, line, etc.)**.
3. Click **Create**. In the **Basic plots (type)** box, select **Scatter**. Select **Price** for the **Y variable**: and **Size** for the **X variable**: Click **Accept**. Click **OK**.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

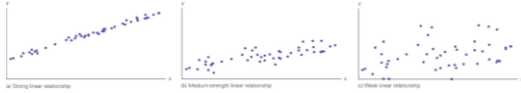
20

20

Chapter 3: Graphical Descriptive Techniques II

Patterns of Scatter Diagrams: Linearity

Scatter Diagrams Depicting linearity



To determine the strength of the linear relationship, we draw a straight line through the points in such a way that the line represents the relationship. If most of the points fall close to the line, we say that there is a linear relationship. If most of the points appear to be scattered randomly with only a semblance of a straight line, there is no, or at best, a weak linear relationship. The three scatter diagrams depicted here exhibit various levels of linearity. Note that there may well be other types of nonlinear relationships, such as quadratic or exponential.

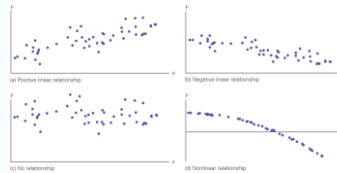
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

21

21

Patterns of Scatter Diagrams: Direction

Scatter Diagrams Describing Direction



If one variable increases when the other does, we say that there is a positive linear relationship.

When the two variables tend to move in opposite directions, we say that there is a negative linear relationship. (The terms positive and negative will be explained in Chapter 4.)

The four scatter diagrams depicted show positive and negative linear relationships, no relationship, and a nonlinear relationship.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

22

22

Interpreting a Strong Linear Relationship

The fact two variables are linearly related does not imply that one causes the other.

In fact, we can never conclude that one variable causes another variable, and express the fact more eloquently as:

Correlation is not causation



Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

23

23

Graphical Excellence

Graphical excellence is achieved when the following characteristics apply:

1. The graph presents large data sets concisely and coherently.
2. The ideas and concepts the statistics practitioner wants to deliver are clearly understood by the viewer.
3. The graph encourages the viewer to compare two or more variables.
4. The display induces the viewer to address the substance of the data and not the form of the graph.
5. There is no distortion of what the data reveal.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

24

24

Chapter 3: Graphical Descriptive Techniques II

- **Data visualization** describes the use of graphical displays to summarize and present information about a data set.
- The goal is to communicate as effectively and clearly as possible the key information about the data.
- Creating effective graphical displays is as much art as it is science.
- Here are some guidelines . . .
 - Give the display a clear and concise title.
 - Keep the display simple.
 - Clearly label each axis and provide the units of measure.
 - If colors are used, make sure they are distinct.
 - If multiple colors or lines are used, provide a legend.

25

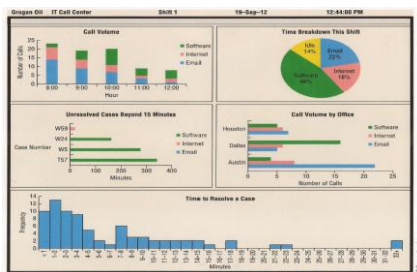
25

- A **data dashboard** is a widely used data visualization tool.
- It organizes and presents **key performance indicators (KPIs)** used to monitor an organization or process.
- It provides timely, summary information that is easy to read, understand, and interpret.
- Some additional guidelines include . . .
 - Minimize the need for screen scrolling.
 - Avoid unnecessary use of color or 3D.
 - Use borders between charts to improve readability.

26

26

Data Dashboard Example

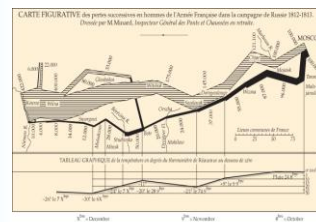


27

27

The Best Chart Ever Drawn

Chart Depicting Napoleon's Invasion and Retreat from Russia in 1812



The striped band represents a time series depicting the size of the French army at various times and places on the map.

When Napoleon invaded Russia in June, he had 422,000 soldiers. By the time the army reached Moscow in October, the number had dwindled to 100,000.

The black band represents the army in retreat. By the time the army returned to France, only 10,000 soldiers remained.

The chart also shows a time series for the temperature, which dipped to -30 °C on December 6.

Source: Charles F. K. Keller, Gerald, Statistics for Management and Economics, 12th Edition, © 2023 Cengage. All Rights Reserved.

28

28

Chapter 3: Graphical Descriptive Techniques II

Graphical Deception

It is important to evaluate critically the information presented by means of graphical techniques. Graphical techniques create a visual impression, which can be easily distorted.

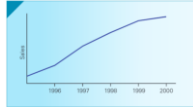
The Canadian Institute of Chartered Accountants began setting guidelines for financial graphics, after a study of annual reports of major corporations found that 8% contained at least one misleading graph to cover up bad results.

It should be noted that misleading graphs may at times be the result of inexperience, rather than deception.

To the right we have a time series showing sales growth without a scale on one axis, making it impossible to assess the growth magnitude.

The next few slides illustrate a few more popular deceptions.

Graph Without Scale



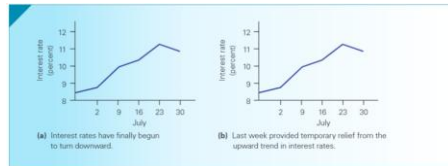
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

29

29

Graphical Deception: Captions

Graphs with Different Captions



Your impression of the trend in interest rates might be different, depending on whether you analyze a graph carrying caption (a) or caption (b).

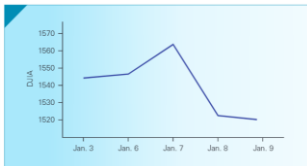
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

30

30

Graphical Deception: Misleading Drop

Graph Showing Drop in the DJIA



Perspective is often distorted if only absolute changes in value, rather than percentage changes, are reported. On January 9, 1986, newspapers reported that the DJIA reported the worst single day loss ever.

However, the 2.5% loss, albeit significant, was nowhere near the Black Friday's decline of 12.8% in 1929.

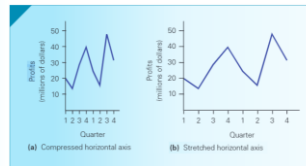
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

31

31

Graphical Deception: Conveying Stability

Graphs Showing Considerable Swings or relative Stability



The opposite effect is obtained by stretching the horizontal axis; that is, spreading out the points on the horizontal axis to increase the distance between them so that slopes and trends will appear to be less steep.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

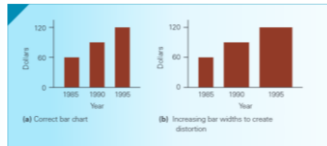
32

32

Chapter 3: Graphical Descriptive Techniques II

Graphical Deception: Distorted Bar Chart

Graphs Showing Correct and Distorted Bar Charts



- The bar chart correctly uses bars of equal width so that both the height and the area of each bar are proportional to the dollar expenditure.
- The growth in expenditures is exaggerated by increasing both the bars' width and height leaving the viewer with the mistaken impression that food expenditures increased far more than they do.

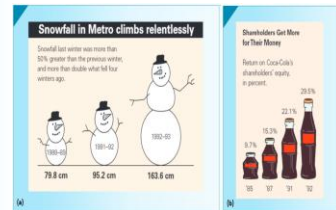
Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

33

33

Graphical Deception: Misuse of Pictograms

Misused and Correct Use of Pictograms



Just like with a bar chart, a pictogram containing objects such as bags of money, people, or animals, can also be misused, exaggerating the growth it is meant to convey.

- Exaggerated use of a pictogram that shows the increase in snowfall over the years depicting a snowman that grows in width as well as height.
- A proper use of a pictogram, in which all the pictures of Coca-Cola bottles are drawn with the same width and varying height.

Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved

34

34

Dot Plot - Hudson Auto Repair

The manager of Hudson Auto would like to have a better understanding of the cost of parts used in the engine tune-ups performed in her shop. She examines 50 customer invoices for tune-ups.

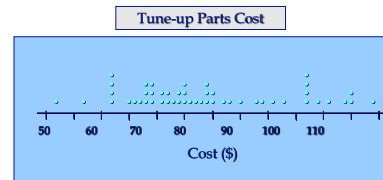
Sample of Parts Cost (\$) for 50 Tune-ups

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

35

35

Dot Plot - A plot of each data value on a scale or number line



- A horizontal axis shows the range of data values.
- Then each data value is represented by a dot placed above the axis.

36

36

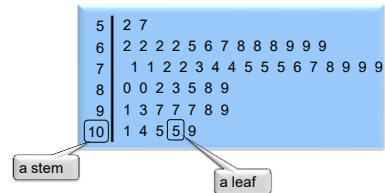
Chapter 3: Graphical Descriptive Techniques II

Stem and Leaf Display

- A stem-and-leaf display shows both the rank order and shape of the distribution of the data.
- It is similar to a histogram on its side, but it has the advantage of showing the actual data values.
- The first digits of each data item are arranged to the left of a vertical line.
- To the right of the vertical line we record the last digit for each item in rank order.
- Each line (row) in the display is referred to as a stem.
- Each digit on a stem is a leaf.

37

- Example: Hudson Auto Repair



38

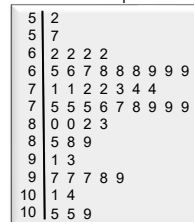
Stretched Stem-and-Leaf Display

- If we believe the original stem-and-leaf display has condensed the data too much, we can stretch the display vertically by using two stems for each leading digit(s).
- Whenever a stem value is stated twice, the first value corresponds to leaf values of 0 - 4, and the second value corresponds to leaf values of 5 - 9.

39

Stretched Stem-and-Leaf Display

- Example: Hudson Auto Repair



40

Chapter 3: Graphical Descriptive Techniques II

Additional – Stem and Leaf Notation

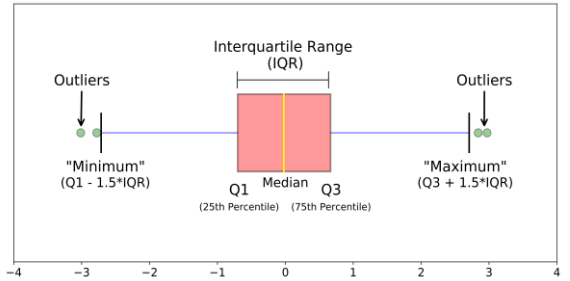
- The Stem-and-Leaf diagram is an alternative way to write out a distribution
- For example, the distribution [12, 14, 19, 23, 27, 29, 31, 33, 36, 38, 44, 45, 48, 55, 59, 60, 67] can be rewritten in Stem-and-Leaf notation as follows:

```
1 2 4 9
2 3 7 9
3 1 3 6 8
4 4 5 8 5 9
5 5 9
6 0 7
```

- Tens Value:** the first number in each row. In the above example, the tens values are 1, 2, 3, 4, 5, 6
- Ones Value:** each number that follows in that row. In the first row, for example, the ones values are 2, 4, 9.

41

How to Read a Box Plot



42

Stacked Bar Chart

- A stacked bar chart is another way to display and compare two variables on the same display.
- It is a bar chart in which each bar is broken into rectangular segments of a different color.
- If percentage frequencies are displayed, all bars will be of the same height (or length), extending to the 100% mark.

[Restaurant.xlsx](#)

43

Stacked Bar Chart



44

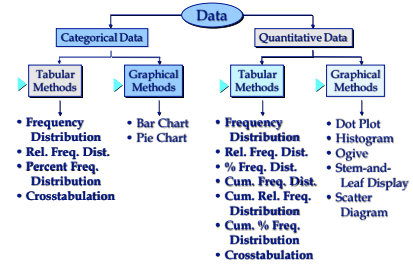
Chapter 3: Graphical Descriptive Techniques II

Choosing the Type of Graphical Display

Bar Chart	to show the frequency distribution or relative frequency distribution for categorical data
Pie Chart	to show the relative frequency or percent frequency for categorical data
Dot Plot	to show the distribution for quantitative data over the entire range of the data
Histogram	to approximate the relationship of data in a scatter diagram
Stem-and-Leaf Display	to show both the rank order and shape of the distribution for quantitative data
Side-by-Side Bar Chart	to compare two variables
Stacked Bar Chart	to compare the relative frequency or percent frequency of two categorical variables
Scatter Diagram	to show the relationship between two quantitative variables
Trendline	to show the frequency distribution for quantitative data over a set of class intervals

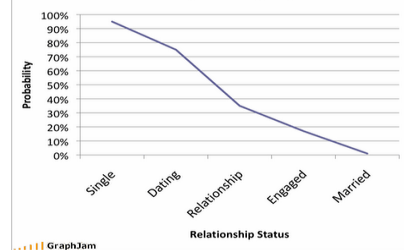
45

Tabular and Graphical Methods



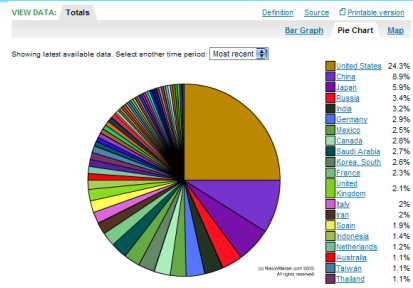
46

Probability of Being Right



<http://www.amstatweb.com/The%20Mag/content/uploads/2013/11/probability-of-being-right-graph.png>

47



Keller, Gerald, Statistics for Management and Economics, 12th Edition. © 2023 Cengage. All Rights Reserved.

48